

A Framework for Supporting User-Centric Collaborative Information Seeking

by
Chirag Shah

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Information & Library Science.

Chapel Hill
2010

Approved by:

Gary Marchionini, Advisor

Deborah Barreau, Committee Member

Susan Dumais, Committee Member

Diane Kelly, Committee Member

Barbara Wildemuth, Committee Member

© 2010
Chirag Shah
ALL RIGHTS RESERVED

Abstract

**CHIRAG SHAH: A Framework for Supporting User-Centric Collaborative
Information Seeking.
(Under the direction of Gary Marchionini.)**

Collaboration is often required or encouraged for activities that are too complex or difficult to deal with for an individual. Many situations involving information seeking also call for people working together. Despite its natural appeal and situational necessity, collaboration in information seeking is an understudied domain. The nature of the available information and its role in our lives have changed significantly, but the methods and tools that are used to access and share that information in collaboration have remained largely unaltered. This dissertation is an attempt to develop a new framework for collaborative information seeking (CIS) with a focus on user-centric system designs. To develop this framework, existing practices for doing collaboration, along with motivations and methods, are studied. This initial investigation and a review of literature are followed by a series of carefully created design studies, helping us develop a prototype CIS system, Coagmento. This system is then used for a large scale laboratory experiment with a focus on studying the role and the impact of awareness in CIS projects. Through this study, it is shown that appropriate support for group awareness can help collaborators be more productive, engaged, and aware in collaboration without burdening them with additional load. Using the lessons derived from the literature as well as the set of studies presented in this dissertation, a novel framework for CIS is proposed. Such a framework could help us develop, study, and evaluate CIS systems with a more comprehensive understanding of various CIS processes, and the users of these systems.

Acknowledgments

I would like to thank all of the following people and organizations for their contributions to my research. Without them, this work would not exist.

- Gene Golovchinsky and Jeremy Pickens at FXPAL, California, for introducing me to the area of collaborative information retrieval during my internship in summer 2007.
- The National Science Foundation, for funding my studies and research with IIS grant # 0812363.
- SILS office staff, for timely processing required paperwork, including the IRB applications, for my user studies.
- SILS library staff, for their kind support and cooperation during my laboratory study that lasted about three months.
- Rob Capra, Katrina Muller, Tessa Sullivan, and Heather Bowden, for helping me pilot my laboratory study.
- All of the participants in my user studies, for investing time and energy in my research.
- My advisor Gary Marchionini, for his constant encouragement and guidance on all of my scholarly endeavors at UNC. Without his kindness, support, and trust in

my work, I could not have been able to even embark upon this journey presented here.

- The rest of my dissertation committee members - Diane Kelly, Barbara Wildemuth, Deborah Barreau, and Susan Dumais, for their highly valuable feedback and guidance throughout the process of this dissertation.
- My mother, father, and sister, for their constant love, kindness, and support, even when they are all in different continents.
- And last, but not the least, my wife Lori, for being absolutely supportive of my research and sacrificing her time with me so that I could work on this dissertation. Even in the darkest hour of toil and frustration, she had faith in me and my efforts. Lori also provided valuable feedback on an early draft of this dissertation, as well as a beta version of Coagmento. Without her, this work could not have been possible.

Table of Contents

Abstract	iii
List of Figures	x
List of Tables	xv
1 Introduction	1
1.1 Research problem and questions	4
1.2 Significance of the proposed research	6
1.3 Summary	8
2 Literature Review	10
2.1 Collaboration	11
2.1.1 Definitions and models of collaboration	11
2.1.2 Principles of collaboration	17
2.1.3 Process of collaboration	19
2.1.4 Limitations of collaboration	22
2.1.5 Collaboration in the context of information science	25
2.2 Information seeking	29
2.3 Information filtering	41
2.4 User and system interaction	47
2.4.1 Degree of involvement	52

2.4.2	Explicitness of collaboration	53
2.5	Social networking	55
2.6	Collaborative Information Seeking (CIS)	67
2.6.1	Space and time aspects of CIS	68
2.6.2	User-source-time configuration for CIS	69
2.6.3	Control, communication, and awareness in a CIS environment	72
2.6.4	Co-browsing or collaborative/social navigation	84
2.6.5	Collaborative IR	87
2.6.6	Realization of a collaborative environment	90
2.7	Evaluation	93
2.8	Conclusion	96
3	Design Investigations and System Development	102
3.1	Preliminary design of Coagmento	103
3.2	Study-1: Personal interviews	107
3.2.1	Instances of collaboration (<i>What</i>)	109
3.2.2	Motivations for collaboration (<i>Why</i>)	111
3.2.3	Methods for collaboration (<i>How</i>)	113
3.2.4	Summary	114
3.3	Study-2: Demonstrative walkthroughs	119
3.4	Study-3: Pilot runs	123
3.4.1	Procedure	123
3.4.2	Data description	125
3.4.3	Analysis	136
3.5	Study-4: Participatory design	139
3.5.1	Coming up with design specifications	140
3.5.2	Feedback on preliminary interface	142

3.6	Improved Coagmento	146
3.7	Conclusion	148
4	Experiments and Results	152
4.1	Background	153
4.2	Method	154
4.2.1	Participants	155
4.2.2	Conditions	157
4.2.3	Sessions	161
4.2.4	Tasks	165
4.2.5	Evaluation	167
4.2.6	Supervising the study	174
4.3	Description of overall user activity	175
4.3.1	Webpages and sources	176
4.3.2	Search queries	177
4.4	Analysis	177
4.4.1	Effectiveness	182
4.4.2	Efficiency	183
4.4.3	Awareness	186
4.4.4	Effort	193
4.4.5	Ease of use and satisfaction	195
4.4.6	Engagement	196
4.5	Conclusion	197
5	Discussion	201
5.1	Instances, motivations, and methods for collaboration	202
5.2	Design and development of Coagmento	203

5.3	Role of awareness in CIS environments	205
5.4	Limitations	208
5.5	A framework for CIS environments	210
5.5.1	A model of collaboration	211
5.5.2	A model of CIS	212
5.6	Future work	219
5.6.1	Extending the studies	220
5.6.2	Theoretical issues	220
	Bibliography	223
	A Definitions	238
	B Study-1 Interview Questionnaire	240
	C Laboratory Study Recruiting	242
	D Laboratory Study Demographic Questionnaire	244
	E Query Re-usage Analysis	246
	F Interaction Effects	251

List of Figures

1.1	Contextual depiction of Collaborative Information Seeking (CIS). . . .	9
2.1	A model showing steps to a meaningful collaboration. (Taylor-Powell, Rossing, and Geran, 1998)	13
2.2	A set-based model of collaboration. An inner set is essential to or sup- ports the outer set.	13
2.3	Distinguishing communication, contribution, coordination, cooperation, and collaboration using different variables. A variable is represented with a bar going minimum to maximum from left to right.	18
2.4	Structure of messy problem solving (Denning and Yaholkovsky, 2008). .	22
2.5	Wilson's model of information seeking (T. D. Wilson, 1997).	32
2.6	A comparison of Ellis's and Kuhlthau's frameworks (T. D. Wilson, 1999).	33
2.7	Wilson's new model incorporating information behavior (T. D. Wilson, 1997).	33
2.8	Four layer model of information seeking centered around information access and organization.	35
2.9	Extension of the four layer model of information access and organization.	37
2.10	A model for Collaborative Information Seeking (CIS) extended from the model in Figure 2.8.	40
2.11	A typical scenario of information seeking in an IR environment.	42
2.12	Content-based information filtering.	44
2.13	Collaborative information filtering.	44
2.14	Saracevic's model of stratified interaction.	50
2.15	Looking at collaboration with the amount of user and system involvement.	54

2.16 Looking at collaboration with the amount of explicitness or intention from user and system.	54
2.17 Looking at collaboration with space and time dimensions. (Twidale and Nichols, 1996).	69
2.18 A classification space for CSCW systems (Rodden, 1991).	70
2.19 General awareness supporting serendipitous communication, division of labor, and focused collaboration (Gaver, 1991).	77
2.20 A search visualization in Ariadne (Twidale et al., 1995).	80
2.21 Screenshot of SearchTogether. (a) integrated messaging, (b) query aware- ness, (c) current results, (d) recommendation queue, (e)(f)(g) search but- tons, (h) page-specific metadata, (i) toolbar, (j) browser. (Morris and Horvitz, 2007).	81
3.1 Timeline for Coagmento design and development, along with the studies.	103
3.2 The main interface and its components in Coagmento.	105
3.3 Toolbar provided on top of the document being viewed with Coagmento.	107
3.4 Collaborative systems/methods organized according to time and space aspects.	116
3.5 Query distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).	126
3.6 Query distribution for the groups.	126
3.7 Query distribution summary.	127
3.8 Viewed documents distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).	128
3.9 Viewed documents distribution for the groups.	128
3.10 Viewed documents summary.	129

3.11	Saved documents distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).	130
3.12	Saved documents distribution for the groups.	131
3.13	Saved documents summary.	131
3.14	Snippets distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).	132
3.15	Snippets distribution for the groups.	133
3.16	Snippets summary.	134
3.17	Chat messages distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).	134
3.18	Chat messages distribution for the groups.	135
3.19	Chat messages summary.	135
3.20	Post-task questionnaire responses.	137
3.21	Exit questionnaire responses.	138
3.22	A screenshot of Coagmento with enhanced views of its toolbar and sidebar.	147
3.23	Snippets window.	149
3.24	Annotations window.	149
4.1	Responses to demographic questions.	156
4.2	Interface for condition-1, baseline.	158
4.3	Interface for condition-2, personal peripheral awareness.	159
4.4	Interface for condition-3, group peripheral awareness.	160
4.5	Coagmento workspace during information seeking part.	162
4.6	Coagmento workspace during information compilation part.	162
4.7	Flow of the sessions.	163
4.8	Participants during a study session.	164
4.9	Supervisor's interface.	175

4.10	Example sidebar showing supervisor’s message.	176
4.11	Source-wise query usage and overlap in aggregation.	181
4.12	Method for statistical analysis of the data.	183
5.1	A modified set-based model of collaboration.	212
5.2	A framework for CIS environments (part 1 of 2).	214
5.3	A framework for CIS environments (part 2 of 2).	215
C.1	Study recruiting page 1 of 2.	242
C.2	Study recruiting page 2 of 2.	243
E.1	Query re-usage in a given session for a given task.	246
E.2	Second session query and subquery re-usage rates from the first session.	248
E.3	Average query and subquery re-usage proportion for a participant with respect to other participants.	248
E.4	Edit Distance among the queries for Task-1. X-axis shows Edit Distance between a pair of queries, and Y-axis shows number of queries.	249
E.5	Edit Distance among the queries for Task-2. X-axis shows Edit Distance between a pair of queries, and Y-axis shows number of queries.	250
F.1	Interaction effects between tasks and sessions for post-task reported awareness questionnaire (measured by an index). No significance dif- ference between conditions.	251
F.2	Interaction effects between tasks and sessions for post-task reported cog- nitive load (measured by an index). No significance difference between conditions.	251
F.3	Interaction effects between sessions for end-session ease of use and satis- faction questionnaire (measured by an index). No significance difference between sessions.	252

F.4 Interaction effects between sessions for end-session engagement questionnaire (measured by an index). No significance difference between session. 252

List of Tables

2.1	Various group activities and examples.	26
2.2	Information seeking strategies (Belkin et al., 1995).	51
2.3	Different scenarios of collaborative information processing.	70
3.1	Pilot runs with Coagmento.	123
3.2	Feedback on Coagmento (version without collaboration support), averaged over eight responses.	144
4.1	Summary of constructs, their definitions, and evaluation measures. I = individual measure. G = group measure.	173
4.2	Hypotheses related to various constructs and conditions.	174
4.3	Top 40 sources for Task-1, economic recession.	178
4.4	Top 40 sources for Task-2, social networking.	179
4.5	Search services used for each session.	180
4.6	Usage of search services used for Task-1.	180
4.7	Usage of search services used for Task-2.	181
4.8	Quality of collected snippets as measured by recall and precision. For one-way ANOVA, between groups $df = 2$ and within groups $df = 39$	183
4.9	Summary statistics of different activities per person. For one-way ANOVA, between groups $df = 2$ and within groups $df = 81$	184
4.10	Summary statistics of webpages overlapped between the collaborators in a given pair. For one-way ANOVA, between groups $df = 2$ and within groups $df = 39$	184
4.11	Index for perceived situational awareness questions. For one-way ANOVA on each question, between groups $df = 2$ and within groups $df = 333$	187

4.12	Summary statistics for responses on perceived awareness questions at the end of each session. For one-way ANOVA on each question, between groups $df = 2$ and within groups $df = 165$	187
4.13	Index for absolute difference between two collaborators' reported numbers for group's status. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$	188
4.14	Average amount of time spent and messages exchanged by each participant while reporting project status. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$	189
4.15	Average difference between reported numbers and real numbers for questions on personal progress. For one-way ANOVA on each response, $df = 2$ and within groups $df = 333$	191
4.16	Average difference between reported numbers and real numbers for questions on group progress. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$	191
4.17	Average number of times workspace accessed by each participant. For one-way ANOVA on the index, $df = 2$ and within groups $df = 81$	192
4.18	Average number of key actions, chat actions, and combined actions taken by each team. For one-way ANOVA on the index, $df = 2$ and within groups $df = 39$	193
4.19	Index for cognitive load questionnaire filled by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$	195
4.20	Index for ease of use and satisfaction questions responded by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$	196
4.21	Index for engagement questions responded by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$	197

4.22	Summary of construct, hypotheses, and related findings. The best performing condition, if any, is given in bold letter.	200
5.1	Differentiating between intent and activeness.	216
E.1	Task and session-wise individual query re-usage statistics.	247

Chapter 1

Introduction

It is natural for us to work with others, as *people are social animals*.¹ There are several good reasons for this natural tendency. For one, sometimes a problem is just too complex for a single individual to tackle. Denning and Yaholkovsky (2008) regard such problems as “messy” or “wicked” and argue that collaboration is essential for resolving such messes. One of the issues with these messy problems is that sometimes people cannot even agree on what the problem is and no solution can make everyone happy (Denning, 2007). However, as we will see later, such a disagreement among the individuals becomes one of the strengths of collaboration. This brings us to another big appeal of collaboration: sometimes the whole can be greater than the sum of its parts (Aristotle, 2 BCb).

Several complex situations call for collaboration. When it comes to accessing or processing some information, it seems that in many situations, multiple people working together will be able to do a *better* job than any one of them individually. For instance, Olson, Olson, Storrsen, and Carter (1993) developed *ShrEdit*, a shared text editor. To

¹It was Aristotle who originally said (2 BCc), “*Hence it is evident that the state is a creation of nature and that man by nature is a political animal.*” The adjective that Aristotle used to describe man in Greek is ‘politikos’ which is where we get the English word political. However, to be ‘politikos’ was to be a member of the ‘polis’ or a citizen. Thus, the meaning of ‘politikos’ more resembles the English word social than political. John Locke, a seventeenth century British philosopher, also stated that “*man is by nature a social animal*”, in regards to human nature.

their surprise, they discovered that the groups working with *ShrEdit* generated *fewer design* ideas, but apparently *better* ones. They believed their tool helped the supported groups keep more focused on the core issues in the emerging design, to waste less time on less important topics, and to capture what was said as they went.

Taking a hypothetical example, imagine a group of students working on a research project for a class. They all have the same goal or information need. Following are some of the scenarios that they can follow.

- Divide up sources to explore among them and then combine and consolidate what they found; this can save them time.
- They can also work independently and then compare the information that overlapped between their findings and evaluate commonality as well as contradictions.
- They can identify tasks for their project and assign different people different roles to play in those tasks. For instance, a couple of group members can take responsibility for digging resources out from the library, another one can set up a meeting with a faculty advisor, and a couple of other members can sort, compare, and consolidate the findings. One of the team members can even be the coordinator.
- More interestingly, a subset of students can look at some information together and discuss on-the-fly; this can allow them to develop a better understanding of the information, refine their goals, and enhance further processing of the new information.

In general, in many problem solving processes, it seems natural to expect that, when people work together in collaboration, they:

1. can accomplish more,
2. benefit from one another's experience and expertise on the given topic, and
3. influence one another and develop a more profound understanding than when they are isolated.

This dissertation is an attempt to understand the *situations* where collaboration among like-minded people can benefit the collaborators, the *tools* and *methods* they use for collaboration, and the ways in which appropriate *support* can be provided for facilitating such collaboration. In the process, a framework is presented that explains collaborative processes in information seeking, and supports the development of better tools and techniques.

The fact that collaboration is not always desired or useful is acknowledged here. Even when collaboration is desired or encouraged, it could induce additional costs that include cognitive load and the cost to coordinate various events and participants. Working in a group may not be beneficial if the participants have conflicts of interest, they do not trust one another, or they do not intend to collaborate (London, 1995). Some of the tradeoffs for costs vs. benefits in collaboration are reviewed in Section 2.1. While the disadvantages and the costs of collaboration are important to acknowledge and study, for the most part in this dissertation it will be assumed that collaboration is intentional and mutually beneficial, and that the participants trust one another.

Considering that collaboration could be a natural choice in many situations, there is a lack of support for users in an information seeking domain to work in collaboration. In general, our understanding of how people work in collaboration on information intensive projects, and the tools that support such activities are inadequate (Shah, 2008). A model of Collaborative Information Seeking (CIS) will be proposed as a part of this dissertation with the emphasis on designing and developing systems and services

that help a set of users with a shared information goal to search and use information collaboratively. In order to validate this model, a set of specific situations that serve as instances of the model are presented. Together these instances formulate an overall research problem, which is presented in the following section, along with specific research questions addressed in this dissertation.

1.1 Research problem and questions

Situation: Two or more people have the same information need and they are working together in the same timeframe² to satisfy the need.

Examples:

- A group of students working on a class project.
- A couple looking to buy a house.
- A patron trying to find some information on a topic with the help of a librarian.
- A set of users watching the same video and commenting on or tagging it.

Difficulty:

There is a lack of specially designed tools to facilitate collaboration among a set of users for information seeking. Most search services and information synthesis tools are designed keeping single user environments in mind. People have been using general-purpose tools such as email and IM to collaborate online (Morris, 2008), and there is an absence of integrated environments that not only support search and communication, but also help the collaborators discover and learn about the information that they may not discover if they work in isolation.

²Note that these users do not need to work synchronously, nor do they have to be co-located. Working definitions for such concepts are provided in Appendix A.

The notion of collaboration itself is not well-understood. There is a lack of mapping between people's situations and motivations to collaborate, and the methods and systems that seamlessly support them.

Example:

Imagine a team of analysts surveying the market for a product. First, we do not want them to repeat each others' efforts, unless it is for verification or training purposes. Secondly, given that we successfully distribute work among them, how do we combine and redistribute the results for their individual analysis? Should we divide the task, execute the requests, and distribute the results based on various aspects of the product? Or should this division be based on users' expertise and role in the project? How do we organize all the information found and evaluate it for commonality and conflicts? How can we facilitate effective communication among the analysts?

Many of these questions are often addressed using ad-hoc methods and work-arounds, such as emails, IM, and meetings, as Morris (2008) showed in her survey of knowledge workers. There is a lack of systematic methods and specialized tools that could not only support these needs, but also enhance the experience of collaboration in information intensive domains.

The goal of the work presented here is to understand how users can effectively collaborate, and based on this understanding to build systems that support an intentional (explicit) and interactive collaboration. In order to achieve this goal the following research questions (RQs) will be addressed.

1. What are the instances, motivations, and methods for people working in collaboration in general, and collaborative information seeking in particular?
2. What are the design challenges for supporting user-centric CIS systems?

3. How to measure costs and benefits in a user-centric CIS environment?

In order to address these research questions, this dissertation will present a series of studies, with the first two questions addressed in Chapter 3 through several design studies, and the last one addressed in Chapter 4 with a laboratory study.

1.2 Significance of the proposed research

In many social situations, it is very common to collaborate. These situations span cultural, gender, and age differences. Large, Beheshti, and Rahman (2002) reported results of a study into gender differences in collaborative web³ searching, conducted in a grade-six classroom of a Canadian elementary school. They demonstrated that their participants very often desired to collaborate on search tasks. Morris (2008) showed, from a survey of 204 knowledge workers, that the majority of them wanted to collaborate. Her further explorations (Morris, 2007b) also proved that collaboration in many situations is vital to the success of the process.

Similarly, when it comes to information seeking, collaboration could be a wise choice. Twidale and Nichols (1996) pointed out a problem - *“The use of library resources is often stereotyped as a solitary activity, with hardly any mention in the substantial library science and information retrieval literature of the social aspects of information systems.”* They argued that introducing support for collaboration into information retrieval systems would help users to learn and use the systems more effectively. Levy and Marshall (1994, p.163) noted that *“...support for communication and collaboration is as important for information-seeing activities, and ... indeed, support for the former*

³Throughout this dissertation, the lowercase ‘web’ will be used instead of the more formal version ‘Web’.

is needed to support the latter.” Based on their extensive study with patent office workers, Hansen and Jarvelin (2005) also concluded that the assumption that information retrieval performance is purely individual needs to be reconsidered. Twidale, Nichols, and Paice (1997) showed that users often desire to collaborate on search tasks and argued that it makes sense to consider browsing as a collaborative process unlike how it is presented by a majority of search engines, i.e., a single-user process. They suggested that a truly user-centered system must acknowledge and support collaborative interactions between users. Morris (2007b) proposed that four features of exploratory search experience - coverage, confidence, exposure, and productivity - could be enhanced by providing explicit support for collaborative search and subsequent sense-making processes.

The question is - are these potential benefits enough to motivate people to collaborate even for the tasks that are typically done in solitude? As discussed before, several works have shown that people want to collaborate in many situations (Morris, 2007a), and that it makes sense to work together to solve certain problems (Denning and Yaholkovsky, 2008). However, either most systems do not support such collaboration explicitly or the users may not realize the value in collaborating. It is very important to identify these situations, inform the users about them, and provide support so the users can reap the benefits of collaborating.

While the issue of collaboration has attained considerable attention lately, much work is needed to address even some of the fundamental issues in this field. This dissertation helps in understanding some of the issues in collaboration for various forms of information seeking and provides indications to enhancing user experience in a collaborative environment. One of such issues in a collaborative environment is the importance of awareness and the ways to provide it. We do not have best practices for implementing and evaluating the impact of awareness in CIS systems. The dissertation will attempt

to address this as one of the central issues. This will enable us to develop better CIS systems by leveraging the usefulness of awareness without burdening the users with unnecessary mental and/or physical load.

The findings of the research presented here will augment our understanding of the way collaboration works in the information seeking domain and general behavior of the users in a collaborative environment. In the course of conducting this research various systems and interfaces will also be developed that can serve as testbeds for further research. In addition to this, a contribution will be made to CIS through developing a method to study collaborative environments.

1.3 Summary

In this chapter a research problem that revolves around investigating instances, motivations, and methods of collaboration was identified. It is acknowledged that (1) while collaboration has its natural benefits, it may not always be useful or desired, and (2) the definition of collaboration presented here assumes it to be an intentional or explicitly specified, interactive, and mutually beneficial process that is carried out (most likely over several sessions) among the collaborators.

Since the focus here is on exploring such processes in the information seeking domain, it is important to lay out this understanding of collaborative information seeking (CIS) in the context of collaboration and information seeking. Such a depiction is given in Figure 1.1. As shown in this scheme, for the purpose of this dissertation, information retrieval (IR) is seen as a subset of information seeking (IS). While IR typically assumes that there exists some information that could satisfy the given information need, IS does not have such an assumption.

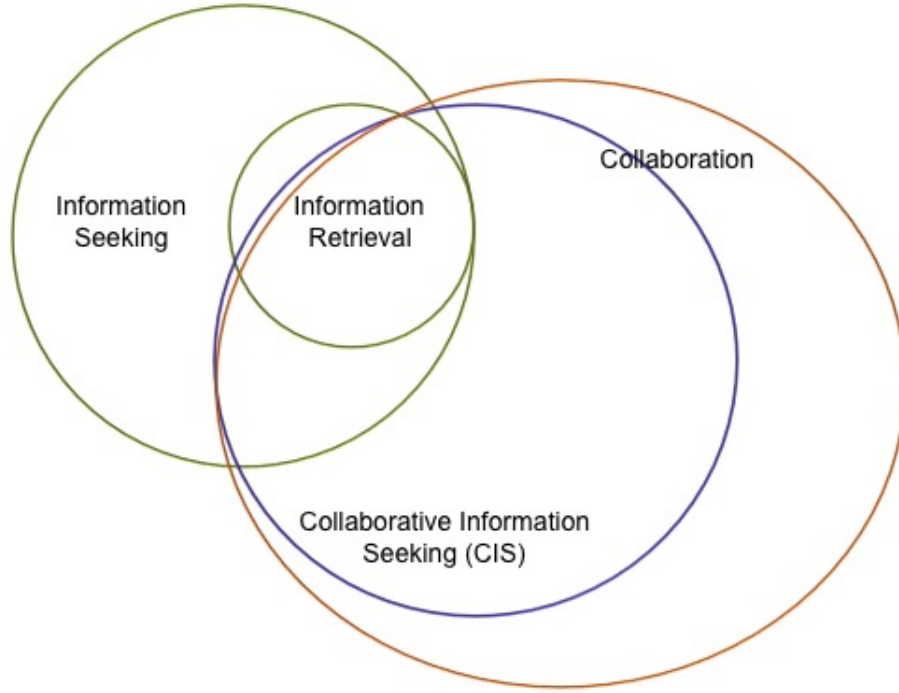


Figure 1.1: Contextual depiction of Collaborative Information Seeking (CIS).

On the other hand, a typical collaboration includes several parts, some of which may be related to information seeking. Thus, in this dissertation, CIS process is seen as a part of a larger context of collaboration. In addition, this CIS is seen as a user-driven (intentional), interactive, and mutually beneficial process.

Keeping in mind this refined definition of CIS, the next chapter will review related domains, such as collaboration, information seeking, information filtering, user and system interaction, and social networking. Using the lessons derived from the literature of related domains, a few key issues will be identified to be addressed in this dissertation. Chapter 3 describes a series of design studies that will help develop an experimental CIS system through several design iterations. Chapter 4 presents a study for testing this system, primarily focusing on its awareness functionalities. A discussion, based on the presented literature review and studies, is given in Chapter 5, concluding this dissertation.

Chapter 2

Literature Review

This work focuses on how people collaborate in information seeking tasks. In addition to this, motivations and costs, as well as the tools and methods used for doing collaboration are reviewed. While the focus here is on studying collaboration in the information seeking process only, it is unlikely that people would collaborate simply for seeking some information without some shared goal. It is highly probable that the process of information seeking has a stage preceding it (e.g., an objective that produces the need for information), and a stage following it (e.g., organization and use of the information collected). In other words, it is difficult to study the act of collaboratively seeking information out of the context in which it is performed. Similarly, it is unrealistic to understand collaboration among users of an information seeking system without characterizing that underlying information seeking process. This realization has led me to study CIS from different points of view, allowing an investigation that starts from a broader aspect of various participants such as collaboration and information seeking, and narrows these views down to a more focused view of CIS.

The CIS related literature reviewed here emanates from five different and broad topics:

1. *Collaboration* - literature covering the general notion of collaboration in various

fields, along with the principles, processes, and limitations of collaboration.

2. *Information seeking* - literature covering a few important models of information seeking and how these models can be mapped to information seeking in collaboration.
3. *Information filtering* - literature relating to filtering information based on the given user's and/or other people's actions and behaviors in the past.
4. *User and system interaction* - literature relating to how humans connect with other humans or systems for interactively seeking information.
5. *Social networking* - literature covering social ties and interactions that could lead to potential stronger connections and collaborations.

2.1 Collaboration

This section is devoted to the discussion and review of the works relating to the general notion of collaboration, and also in the context of information science. Various definitions of collaboration, as well as its relationship with cooperation and coordination are first presented. Next, some works from the literature that discuss the principles of collaboration, the actual process of collaboration used, and the limitations of collaboration are reviewed.

2.1.1 Definitions and models of collaboration

We seem to have an intuitive understanding of what it means to collaborate. As its Latin roots 'com' and 'laborate' suggest, *collaboration* indicates, "to work together". London (1995, p.8) interpreted this meaning as *working together synergistically*. Gray (1989, p.11) defined collaboration as "*a process of joint decision-making among key*

stakeholders of a problem domain about the future of that domain.” Roberts and Bradley (1991, p.209) called collaboration “*an interactive process having a shared transmutational purpose.*”

We often find people using the term ‘collaboration’ in various contexts and interchangeably with terms such as ‘coordination’ and ‘cooperation’. It is very important that we first ground the meaning of the term ‘collaboration’ before addressing various issues regarding doing collaboration. Denning and Yaholkovsky (2008) suggested that coordination and cooperation are weaker forms of working together, and that all of these activities require sharing some information with each other. Taylor-Powell, Rossing, and Geran (1998) added another component to this - contribution, as they realized that in order to have an effective collaboration, each member of the group should make an individual contribution to the collaborative. Using communication, contribution, coordination, and cooperation as essential steps toward collaboration, they showed how a true collaboration requires a tighter form of integration (Figure 2.1).¹

Based on these two works, a model of collaboration is synthesized and presented in Figure 2.2. This model has five sets: *communication* (information exchange), *contribution*, *coordination*, *cooperation*, and *collaboration*. Considering notions of sets, the model shows which activity is supporting which other. For instance, coordination is a subset of collaboration, which indicates that, for a meaningful collaboration, we need to have some way of coordinating people and events. Collaboration is a superset of cooperation, which means in order to have a true collaboration, we need something more than cooperation. These five sets are described below in more details. To help explain these concepts, various activities in the context of a library will also be listed.

¹Available at http://www.empowerment.state.ia.us/files/annual_reports/2001/Collaboration.pdf.

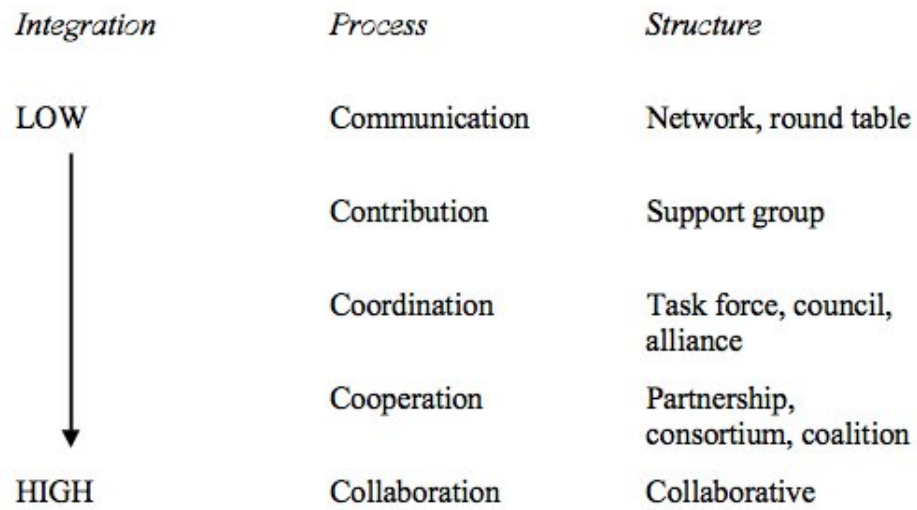


Figure 2.1: A model showing steps to a meaningful collaboration. (Taylor-Powell, Rossing, and Geran, 1998)

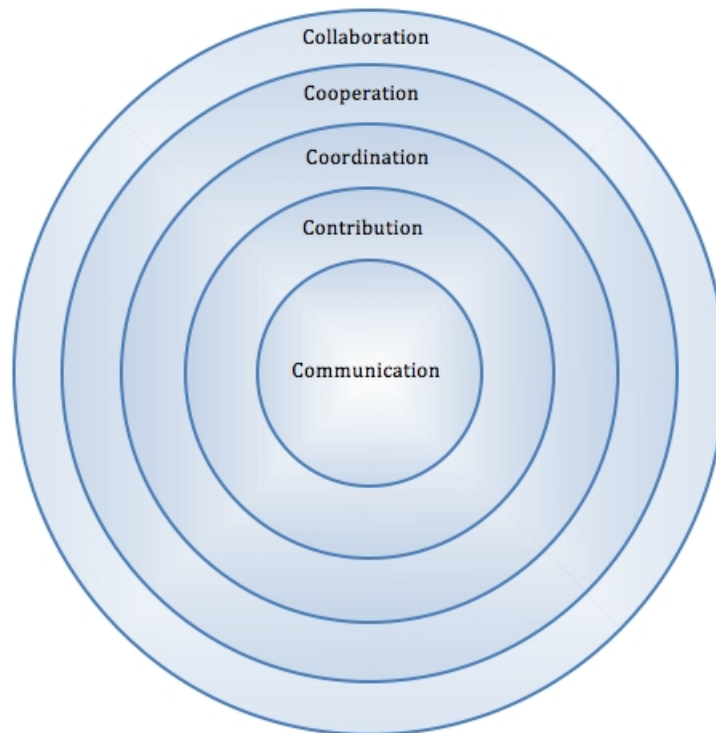


Figure 2.2: A set-based model of collaboration. An inner set is essential to or supports the outer set.

- *Communication.* This is a process of sending or exchanging information, which is one of the core requirements for carrying out collaboration, or maintaining any kind of productive relationship for that matter. For instance, there is a message on the local public library's bulletin board about a book sale for charity the coming weekend. This is a way for the library to communicate with the visitors, which may result in some coordinated event.
- *Contribution.* This is an informal relationship by which individuals help each other in achieving their personal goals. For instance, Mark has some old books that he no longer needs, but they could be of use to others. He, therefore, asks the library if they would take those books. Upon the library's approval (communication), Mark donates the books to the library.
- *Coordination.* This is a process of connecting different agents together for a harmonious action. This often involves bringing people or systems under an umbrella at the same time and place. During this process, the involved agents may share resources, responsibilities, and goals. For instance, Mark decides to study with his fellow student Richard in the library. They both get together at the library on a set day and study their own material. They both are in a way helping each other by keeping each other company (contribution), but they do not share a specific goal that they are working toward together.
- *Cooperation.* This is a relationship in which different agents with similar interests take part in planning activities, negotiating roles, and sharing resources to achieve joint goals. In addition to coordination, cooperation involves all the agents following some rules of interaction. For instance, if the reference librarian, Carol, had simply pointed Mark to the relevant section for his query, she has cooperated with him, but they did not collaborate. What extends several cooperative acts to

collaboration is an active session of interaction in which both the parties worked together to solve a problem.

- *Collaboration.* This is a process involving various agents that may see different aspects of a problem. They engage in a process through which they can go beyond their own individual expertise and vision by constructively exploring their differences and searching for common solutions. In contrast to cooperation, collaboration involves creating a solution that is more than merely the sum of each party's contribution. The authority in such a process is vested in the collaborative rather than in an individual entity.

For instance, Mark walks up to a reference librarian, Carol, in a public library. He is studying 20th century American poets and trying to find all the material that could be relevant. Carol helps him to locate the section of the library, where Mark can find the books on poetry. However, that would not be enough to be considered collaboration. Mark describes the scope of his study and that gives Carol information to start suggesting specific books to Mark. After skimming through those books over a couple of day's time, Mark thinks he has a lot more material than he needed. He returns to Carol with this concern, and she suggests that he look at a book that is an anthology of American poets. This book not only helps Mark directly, but also lets Carol refine her suggestions as she also flips through it. Finally, Mark has good material on this topic that is not too general and covers most of the information that he would like to have.

Working on this collaborative task required both Mark and Carol to interact with each other (communication) person-to-person at a certain place and time (coordination). The starting point of this collaboration was Carol agreeing to help Mark (cooperation). While Carol seems to have done most of the work (contribution), she also learned a certain number of things in the process. Finally,

while the rewards of this process were different for both the parties, it is important to note that they both worked together to solve the same problem.

In addition to this, we can hope that this solution is also a better one, since often a group of entities are found to create a much better solution than any individual entity by itself (Surowiecki, 2004). Chrislip and Larson (1994, p.5) defined collaboration as a “*mutually beneficial relationship between two or more parties [agents] who work toward common goals by sharing responsibility, authority, and accountability for achieving results.*” Similarly, according to Gray (1989, p.5), collaboration is “*a process through which parties [agents] who see different aspects of a problem can constructively explore their differences and search for solutions that go beyond their own limited vision of what is possible.*”

The difference among these five activities can be summarized using the following variables, which are also depicted in Figure 2.3.

1. *Interaction.* While communication is at the center of other activities, it is possible to have a very little amount of interaction while simply communicating. For instance, a system administrator sending an email to a user for his overused disk quota may not require any further interaction. The administrator may not even personally know or see the user. A typical collaboration, on the other hand, requires a high level of interaction among the participants.
2. *Intent.* Similar to interaction, a collaborative project requires much stronger intent compared to those tasks that are merely coordinating events or one entity cooperating with another.
3. *Trust.* In order to have an effective and mutually beneficial collaboration, the participants need to establish a good amount of trust among them. Such is not a requirement for simply coordinating or cooperating.

4. *Human involvement.* A process of communication may not require much human involvement. For instance, posting a message on a noticeboard is an act of communication, but it does not require the poster or the readers of that message being involved with each other. Collaboration, on the other hand, requires the participants to be actively engaged in the project at hand.
5. *Symmetry of benefits.* The kind of collaboration considered here is by definition mutually beneficial. Thus, it benefits everyone involved in the process. The amount of benefit may vary depending on participants' roles and responsibilities. In contrast, in a cooperative process, one party helps the other party in fulfilling their goals. A student gathering relevant literature for his supervisor is an act of cooperation, contributing to the supervisor's goal. Co-authorship, on the other hand, is an act of collaboration, benefitting all the participants.
6. *Level of awareness.* For an interactive, intentional, and mutually beneficial collaboration to be successful, it is imperative that all the participants be aware of each others actions and contributions. This also helps in establishing a level of trust among the participants. Provision of such awareness may not be a requirement for coordination or cooperation.

2.1.2 Principles of collaboration

Let us now look at the principles or conditions for an effective collaboration. Most researchers agree that an effective collaboration must be *democratic* and *inclusive*, that is, it must be free of hierarchies of any kind and it must include all parties who have a stake in the problem (London, 1995). Note that such democratic collaboration may still have assymetric roles for different collaborators.

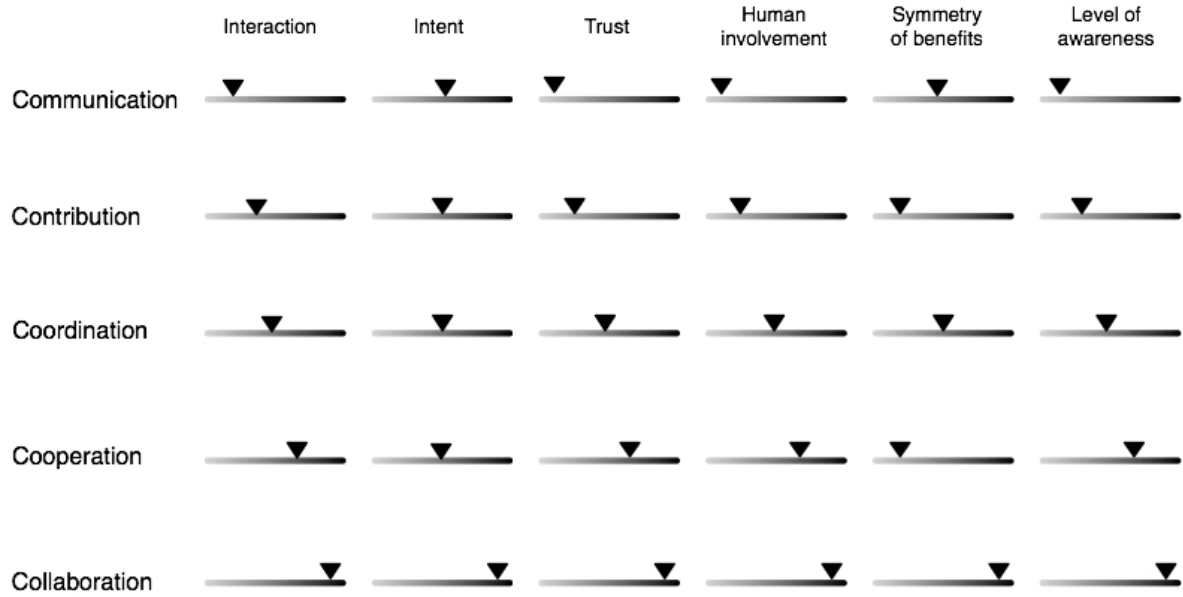


Figure 2.3: Distinguishing communication, contribution, coordination, cooperation, and collaboration using different variables. A variable is represented with a bar going minimum to maximum from left to right.

Regarding *democracy* in collaboration, Flora, Flora, and Fey (2004, p.273) pointed out, “*without community empowerment and broad participation in agenda setting, the decision-making process of discussion, debate, and compromise is relatively meaningless.*” Osborne and Gaebler (1992, p.269) also expressed their views against hierarchies by noting that centralized and hierarchical associations tend to be divided up into many layers and boxes. They recognized a difficulty in carrying out communication across units and between layers, thus inhibiting the real potential of collaboration.

There is also a general agreement about *inclusiveness* in collaboration. Theobald (1987, p.118), for instance, argued that all the leadership of a community must be involved, whether participants fit traditional definitions of leaders or not. Chrislip and Larson (1994, p.52) concurred, reporting that all the successful collaborations they studied involved many participants from several factors, such as government, business, and community groups, as opposed to few participants predominantly from one sector.

Gray (1989) further claimed that collaboration could only be meaningful if the stakeholders are interdependent. She explained, “*collaboration establishes a give and take among the stakeholders that is designed to produce solutions that none of them working independently could achieve.*”

To spell out what situations could create a meaningful collaboration, Surowiecki (2004) presented four conditions for a successful collaboration.

1. *Diversity of opinion.* Each person should have some private information, even if it is just an eccentric interpretation of known facts.
2. *Independence.* People’s opinions are not determined by the opinions of those around them.
3. *Decentralization.* People are able to specialize and draw on local knowledge.
4. *Aggregation.* Some mechanism exists for turning private judgments into a collective decision.

Collaboration, in many situations, is a process that ties people of varying opinions and abilities together. However, the process may not necessarily lead to agreement on all issues. For instance, Gray (1989, p.25) acknowledged that not all collaborations lead to consensus, but added that, when agreements for action are reached, they are always done so by consensus. Denning and Yaholkovsky (2008) also noted that it is solidarity, not software, that generates collaboration.

2.1.3 Process of collaboration

Following is a list of points compiled by London (1995) from a combination of sources, including Chrislip and Larson (1994), Straus and Williams (1986), Portnoy (1986), and Gray (1989), that one needs to consider before starting a collaborative process.

1. What are the structural relationships between the parties and the possible power issues inherent in the collaborative arrangement?
2. Is there a clear understanding among all the parties of the respective goals of the other participants?
3. What form of leadership is required to facilitate the process?
4. Does the project have some form of integrating structure, such as a cross-section of steering committees, to facilitate and coordinate decision-making and implementation?
5. Will the project be more effective with a neutral, third party mediator?
6. Should the media be involved?
7. Does the project have enough time, money, and staff support?

Such questions and considerations are useful to understand as collaboration may induce costs, and we may want to make sure that the given problem and circumstances do call for collaboration.

Let us now look at what collaboration involves. Collaboration is typically a complex process involving a number of phases, a variety of interactions, and other sub-processes. Gray (1989) identified three major phases of collaboration.

1. *Pre-negotiation or problem-setting phase.* This phase is often the most difficult, and involves six issues to be addressed.
 - (a) The parties must arrive at a shared definition of the problem, including how it relates to the interdependence of the various stakeholders.
 - (b) The parties must make a commitment to collaborate.

- (c) Other stakeholders must be identified whose involvement may be necessary for the success of the endeavor.
 - (d) The parties must acknowledge and accept the legitimacy of the other participants.
 - (e) The parties must decide what type of convener or leader can bring the parties together.
 - (f) The parties must determine what resources are needed for the collaboration to proceed.
2. *Direction-setting phase.* During this phase, the parties need to identify the interests, which brought them together, determine how they differ from the interests of the others, set directions, and establish shared goals. This phase is characterized by six steps.
- (a) Establishing ground rules
 - (b) Setting the agenda
 - (c) Organizing sub-groups (especially if the number of issues to be discussed is large or the number of stakeholders exceeds the twelve to fifteen member limit for effective group functioning)
 - (d) Undertaking a joint information search to establish and consider the essential facts of the issue involved
 - (e) Exploring the pros and cons of various alternatives
 - (f) Reaching agreement and settling for a course of action
3. *Implementation phase.* During this final phase, the participants go through the following steps.
- (a) Participating groups or organizations deal with their constituencies.

- (b) Parties garner the support of those who will be charged with implementing the agreement.
- (c) Structures for implementation are established.
- (d) The agreement is monitored and compliance is ensured.

Similar to these three phases of Gray, Denning and Yaholkovsky (2008) provided three main stages of solving a complex problem: design, collaboration, and follow-through (Figure 2.4). The authors defined the collaboration stage as a set of five specific sub-stages: (1) declare, (2) connect, (3) listen to and learn all perspectives, (4) allow a “we” to develop, and (5) create together.

Individual collaborative process will depend on various factors such as the nature of collaboration, the number of people or parties involved, the timeframe, and the resources at hand.

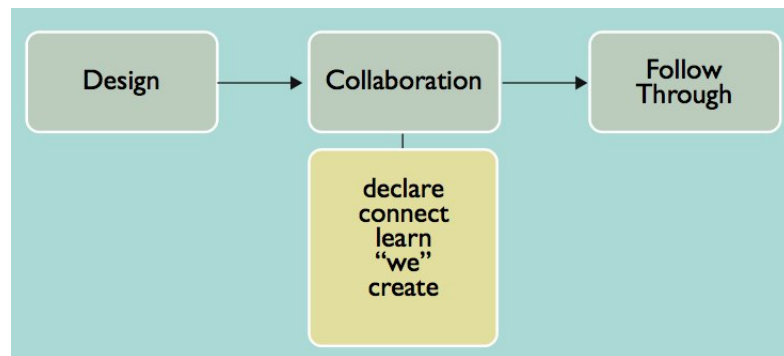


Figure 2.4: Structure of messy problem solving (Denning and Yaholkovsky, 2008).

2.1.4 Limitations of collaboration

It was noted earlier that in many situations, collaboration is a natural choice, especially for solving the problems that are hard (Denning and Yaholkovsky, 2008). However, one must also need to understand the costs and benefits associated with a collaborative process in order to evaluate the usefulness and the effectiveness of that collaboration.

London (1995) identified the following limitations of a collaborative process.

1. Collaboration is a notoriously time-consuming process and is not suitable for problems that require quick and decisive action.
2. Power inequalities among the parties can derail the process.
3. The norms of consensus and joint decision-making sometimes require that the common good take precedence over the interests of a few.
4. Collaboration works best in small groups and often breaks down in groups that are too large.
5. Collaboration is meaningless without the power to implement final decisions.

Gray (1989) listed five circumstances under which it is best not to collaborate:² (1) when one party has unchallenged power to influence the final outcome; (2) when the conflict is rooted in deep-seated ideological differences; (3) when the power is unevenly distributed; and (4) when constitutional issues are involved or legal precedents are sought, and (5) when a legitimate convener cannot be found.

Sometimes we see collaboration forced on a group of people. Examples of such forced collaborations include the merger of two companies or instructor-enforced class groups. In such situations, the collaboration may start with acts of cooperation, where the participants are merely following a set of rules working with others in the group. Later, such cooperative events may result in collaboration as the participants take active part (intention) in driving the process of working together for a common goal. However, collaboration may still not be successful if the participants do not trust each other or if the power and benefits are unbalanced (London, 1995; Gray, 1989).

²The author's observations and claims are based on her studies in the civic collaboration domain.

Collaboration can also have limited advantages if the costs and benefits are unevenly distributed among the participants. As one of the eight challenges of groupware system development, Grudin (1994) talked about disparities in benefits and responsibilities among the participants. He claimed that it is almost impossible to have a groupware system in which every participant does the same amount of work and/or benefits the same. His examples show how some participants of a groupware system have to do additional work while not getting equivalent benefits. Due to such inequality, the groupware application may become less and less useful and may even stop being used.

While the kind of collaboration that is considered here (intentional and mutually beneficial) is slightly different than Grudin's notion of groupware, and the discussed CIS systems are considerably different than the groupware systems Grudin talked about, several of the issues he raised and the recommendations he made are relevant.

For the above-mentioned challenge, Grudin's recommendation for the system developer was to make sure that the system benefits all the participants. At the same time, he identified the difficulties in doing so, because, while the managers or higher authorities are gaining more benefits, they are the decision makers. Pleasing the upper management personnel is equally important, if not more, as pleasing the rest of the participants who have to do additional work.

Part of the problem in disparity of benefits is also due to the highly asymmetric roles in such kind of collaborations. Having diversity in collaboration could be very useful for a successful collaboration (Surowiecki, 2004), but as Aneiros and Estivill-Castro (2003) argued, roles according to positions (manager vs. knowledge workers), can create several constraints while seeking information collaboratively. They advised against such a master/slave model of collaboration and proposed a way of unconstrained co-browsing with asymmetric roles.

2.1.5 Collaboration in the context of information science

To understand the model of collaboration presented earlier (Figure 2.2) in the context of information science, these five sets are listed in Table 2.1 with examples. Sending an email or conversing on an IM client is a form of communication. Of course, this communication could be a part of a collaborative project (see that communication is a subset of collaboration in Figure 2.2.). In fact, email is one of the most used methods of communication in a collaborative project (Morris, 2008). While communication tools can be used to send one agent's contribution to another, there are specialized tools and places for doing so. Among these, online support groups and social QandA sites, such as Yahoo! Answers, are very popular. The asker and answerer (contributor) on these sites are not truly collaborating; one agent (user) is merely helping the other with his information need. To make such a process more effective and explicit, people use conference calls or net meetings. This requires coordinating the agents (people as well as systems). Once again, such a coordinated event could be a part of a collaborative project. If we take coordinated contribution with a set of rules that the participating agents need to follow, we can find examples of cooperation. On Wikipedia, the participants not only contribute in a coordinated fashion, but there are also rules for this participation and contribution that the users need to follow. In case of a disagreement, there are guidelines that suggest how to make this interaction work. Beyond cooperative activities, collaboration involves a group of agents working toward a common goal with explicit interactions. Imagine co-authoring an article. The authors involved in this project not only contribute and coordinate with other authors, but they also follow some sorts of rules that guide the aggregation of contributions and their mutual interactions. The authors also interact with each other to create this common product, which may be greater than the summation of their individual contributions.

Let us look back to the terms 'coordination' and 'cooperation', and see how they

Table 2.1: Various group activities and examples.

Activity	Definition	Examples
Communication	Exchanging information between two agents	Email, chat
Contribution	Offering of an individual agent to others	Online support groups, social QandA
Coordination	Connecting different agents in a harmonious action	Conference call, net meeting
Cooperation	Agents following some rules of interaction	Wikipedia, Second Life
Collaboration	Working together synergistically to achieve a common goal	Brainstorming, co-authorship

fit around this understanding of collaboration. Austin and Baldwin (1991) noted that while there are obvious similarities between cooperation and collaboration, the former involves pre-established interests, while the latter involves collectively defined goals. Malone (1988, p.5) defined coordination as *“the additional information processing performed when multiple, connected actors pursue goals that a single actor pursuing the same goals would not perform.”* While this definition is close to the one we have seen about collaboration, one can argue that it still fits in the model described in Figure 2.2 since it says nothing about creating solutions. For instance, organizing a meeting involves coordinating among the attendees, but it is not a collaborating activity.

From the definitions and models described above, we can conclude that, in order to have a successful collaboration while seeking information, we need to create a supportive environment where:

1. The participants of a team come with different backgrounds and expertise.
2. The participants have opportunities to explore information on their own without being influenced by the others, at least during a portion of the whole information seeking process.

3. The participants should be able to evaluate the discovered information without always consulting others in the group.
4. There has to be a way to aggregate individual contributions to arrive at the collective goal.

One important aspect of the above requirements that is missing is the kind of task. There may not be much point in collaborating for simple fact-finding information tasks. As Morris and Horvitz (2007) expressed, tasks that are exploratory in nature are likely to benefit from collaboration.

How the concept of collaboration relates to CIS

The notion of collaboration and the requirements to have a successful collaboration presented in this section will now be taken to propose (1) a set of conditions under which collaboratively seeking information is useful, and (2) a set of guidelines for building a successful CIS environment.

The conditions under which collaboratively seeking information is useful are given below. They are not very different from those of any other kind of collaborative process.

1. Common goal and/or mutual benefits

This is covered in the definition of the kind of collaboration that is under consideration here. Often, it is the common goal and/or the possibility of mutual benefits that brings people together for collaboration. For the most part, this is not a function of a system. A system can provide support for people with common goals who want to collaborate and reap the benefits of that collaboration, but does not typically initiate the collaboration. On the other hand, a few systems provide a functionality of connecting the visitors to the same websites in order for them to have a possible collaboration, such as the one given by Donath and

Robertson (1994). These systems are based on the assumptions that the people browsing the same websites may have the same information needs.

2. *Complex task*

Morris and Horvitz (2007) showed that there are not many benefits for collaborating on simple tasks, such as fact-finding. Denning and Yaholkovsky (2008) also recognized the benefit of collaborating while solving “messy” or “wicked” problems. While listing the conditions under which it is not useful to collaborate, London (1995) argued that if a task is simple enough, there is no point in collaborating.

This may imply that the task should be exploratory in nature, and may span several sessions.

3. *High benefits to overload ratio*

Often, a simple divide and conquer strategy could make collaboration successful. However, such a process may have its overhead. London (1995) noted that collaboration is only useful if such an overhead is acceptable for the given situation. Fidel, Pejtersen, Cleal, and Bruce (2004) showed that collaboration induces additional cognitive load, what they referred to as the *collaborative load*. The collaboration in question has to meet or exceed the benefits expectations for it to be viable with the cognitive load that it brings.

4. *Insufficient knowledge or skills*

A common reason to collaborate is the insufficient knowledge or skills an individual possesses for solving a complex problem. In such cases, the participants can collaborate so that they can achieve something bigger or better than what they each could do individually. In other words, the whole can be bigger than the sum of all.

The guidelines for building a successful CIS environment, following the discussion of the model in Figure 2.2, and derived from the discussion in this section, are given below.

1. A CIS system should provide effective ways for the participants to communicate with each other.
2. A CIS system should allow (and encourage) each participant to make individual contributions to the collaborative.
3. A CIS system should coordinate participant actions, information requests, and responses to have an active and interactive collaboration. This collaboration could be synchronous or asynchronous, and co-located or remote.
4. Participants need to agree to and follow a set of rules to carry out a productive collaboration. For instance, if they have a disagreement on the relevancy of an information object, they should discuss and negotiate; they should arrive at a mutually agreeable solution rather than continuing to dispute it. The system needs to support such a discussion and negotiation process among the participants.
5. A CIS system should provide a mechanism to let the participants not only explore their individual differences, but also negotiate roles and responsibilities. There may be a situation in which one participant leads the group and others follow (cooperate), but the real strength of collaboration lies in having the authority vested in the collective.

2.2 Information seeking

Marchionini (1995) defines information seeking as a process in which humans purposefully engage to change their state of knowledge. This process of information seeking

goes beyond simply retrieving information; it is usually associated with higher level cognitive processes such as learning and problem solving (Marchionini, 1989). Dervin and Nilan (1986) presented a view of information seeking that emphasized communication and the needs, characteristics, and actions of the information seeker rather than mere representation, storage, and retrieval of information.

Several models have been proposed to understand and explain the information seeking process and information seeking behavior. Dervin (1977) presented a model with three phases of users facing and solving their information problems.

1. *Situation*. This phase establishes the context for the information need, called situation.
2. *Gap*. People often find that given the situation, there is a gap between what they understand and what they need to make sense of the current situation.
3. *Use*. Once this gap is realized, it is manifested by questions. The answers to these questions are put to use and then the user moves on to the next situation.

Belkin (1980) proposed a model of information seeking that focuses on information seekers' anomalous states of knowledge (ASK). In this model, the information seekers do not have a clear understanding of the problem they are trying to solve or the information needed to do so. Information seekers have to go through the stage of articulating their search request, and the search system helps to refine that request. Thus, the ASK model serves as a theoretical basis for the design of interactive information systems.

Wilson (1980) presented a model of information seeking process, which he extended (Figure 2.5) showing how the work by Ellis (1989) can be incorporated into the original model. The stages of information seeking behavior shown in this figure were reported in (Ellis, Cox, and Hall, 1993) as *Starting*, *Chaining* (following citation linkages), *Browsing*, *Differentiating*, *Monitoring*, *Extracting*, *Verifying*, and *Ending*. Ellis' model is

based on empirical research and has been tested in various domains, most recently in the context of an engineering company (Ellis and Haugan, 1997). Kuhlthau, in her work (Kuhlthau, 1991) and (Kuhlthau, 1994), supplemented Ellis' work by attaching to the stages what she called *information seeking process*, the associated feelings, thoughts and actions, and the appropriate information tasks. Her model's stages were: *Initiation, Selection, Exploration, Formulation, Collection and Presentation*. Wilson (1999) presented a comparison of Ellis' and Kuhlthau's models (Figure 2.6): "*...the two models are fundamentally opposed in the minds of the authors: Kuhlthau posits stages on the basis of her analysis of behaviour, while Ellis suggests that the sequences of behavioural characteristics may vary.*"

Based on these works of information seeking, Wilson revised his model (Wilson, 1997) and extended it to incorporate information behavior (Figure 2.7). He proposed to fill the *gap* that Dervin labelled in her model by inserting a concept of activating mechanism. Wilson proposed to do this using the stress/coping model proposed by Miller and Mangan (1983).

Using the work by Belkin, Dervin, Ellis, and Kuhlthau as reported above, Westbrook (1993) proposed a model that included stages such as "Need". Her full set is: *Needing, Starting, Working, Deciding and Closing*. Marchionini (1995) presented an information seeking model where the process of information seeking is composed of eight subprocesses which develop in parallel: (1) recognize and accept an information problem, (2) define and understand the problem, (3) choose a search system, (4) formulate a query, (5) execute search, (6) examine results, (7) extract information, and (8) reflect/iterate/stop.

How information seeking relates to CIS

The models described above assume individual information seekers. In this section,

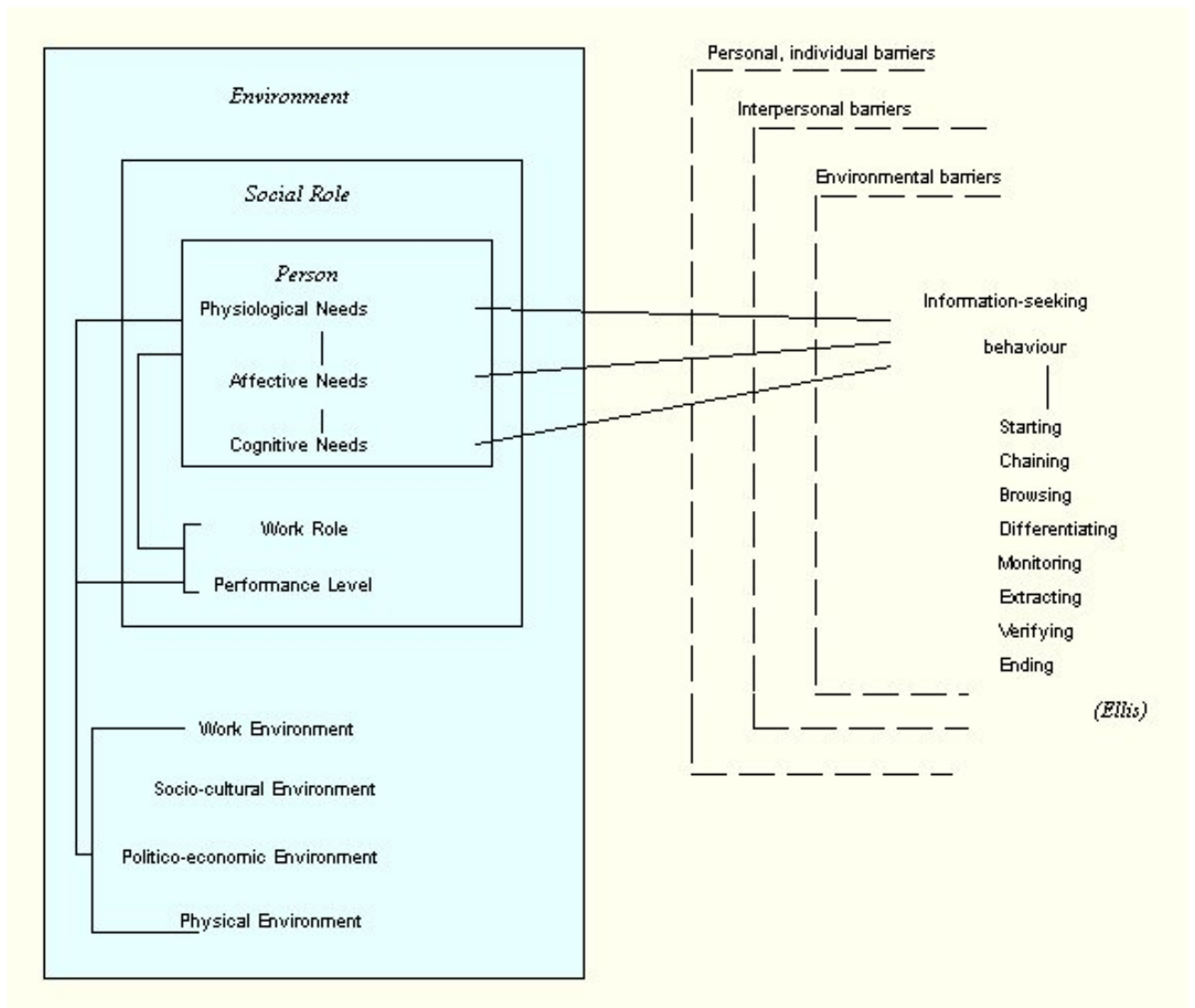


Figure 2.5: Wilson's model of information seeking (T. D. Wilson, 1997).

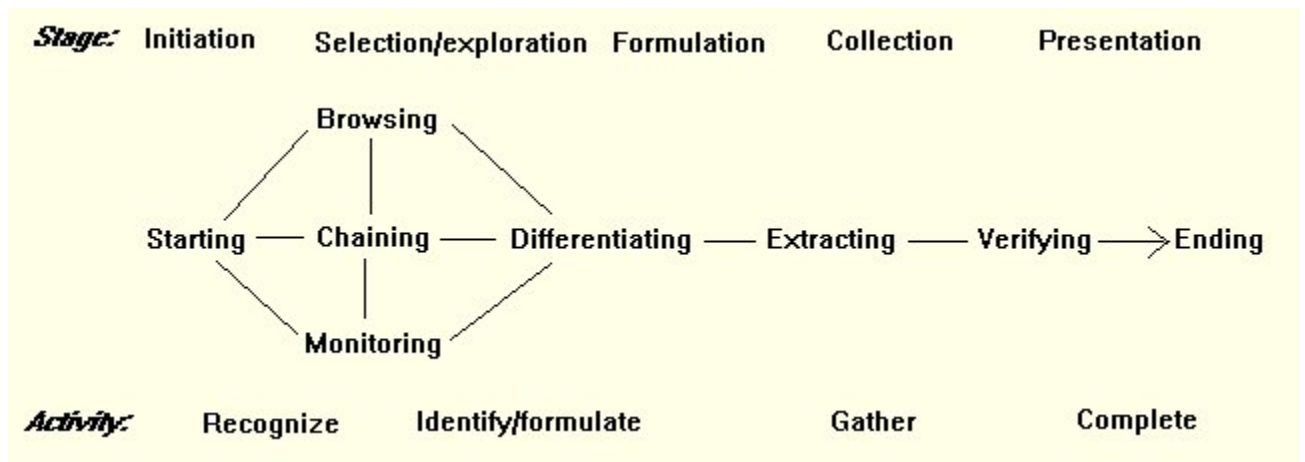


Figure 2.6: A comparison of Ellis's and Kuhlthau's frameworks (T. D. Wilson, 1999).

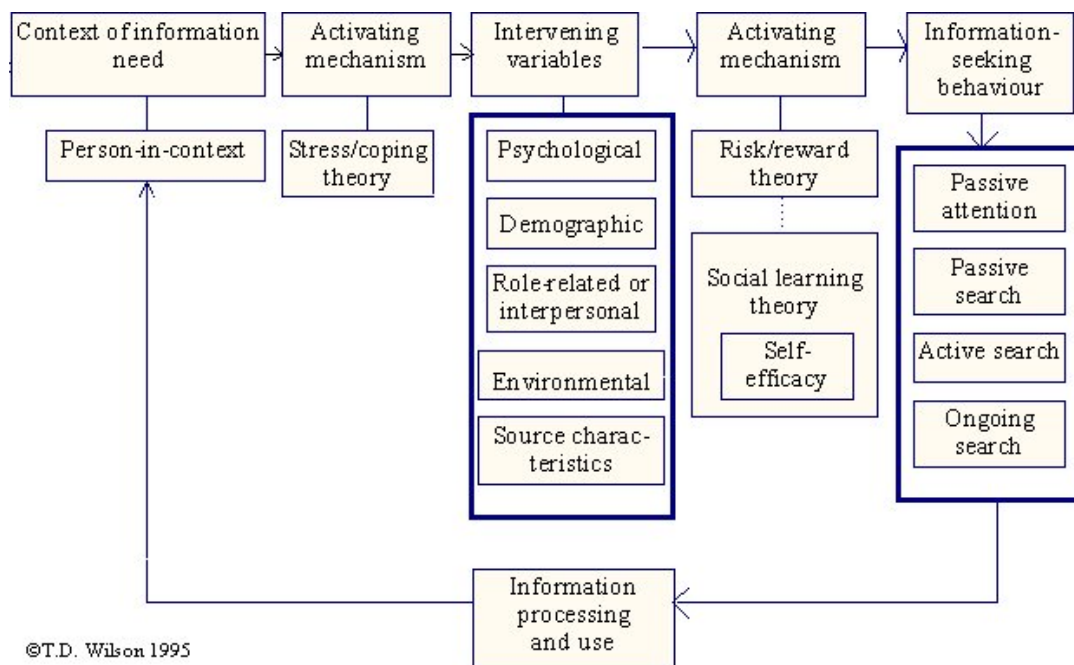


Figure 2.7: Wilson's new model incorporating information behavior (T. D. Wilson, 1997).

extensions to CIS are outlined. As we can see, all of these models of information seeking incorporate various stages of an information seeker's behavior. For the purpose of this dissertation, the focus is on how an information seeker can use various methods and tools to seek information that is mediated by computer(s), and what he³ does when/if information that he is looking for is found. Of course, the information that he is looking for may not exist and the seeker may not have a clear idea about the kind of information that may be useful for his task. Marchionini's (1995) model incorporating eight subprocesses of information seeking can be useful for the purpose of defining such a process-based model for CIS. In order to map various processes in a CIS environment to Marchionini's model, the focus will be on how two participants can seek and process information in collaboration.

A four layer model of information seeking, centered around information access and organization is presented in Figure 2.8 to facilitate the discussion. On the left side, four layers are labeled; on the right side, examples are given for these layers; and in the middle, a typical scenario is presented. These four layers are described below in detail.

Layer-1: Sources

This layer contains information in various sources and formats (structured, semi-structured, and unstructured). The sources include digital libraries, wikis, blogs, databases, and webpages; formats include text, images, and videos. In fact, a person or a group of people could also act as a source.

Layer-2: Tools

This layer consists of tools and techniques a user can use to interface with the sources and access the information of layer-1. They include search services, relevance feedback mechanisms (Buckley, Salton, and Allen, 1994), and query term suggestions (Anick,

³Throughout this dissertation, 'he' implies 'he/she'.

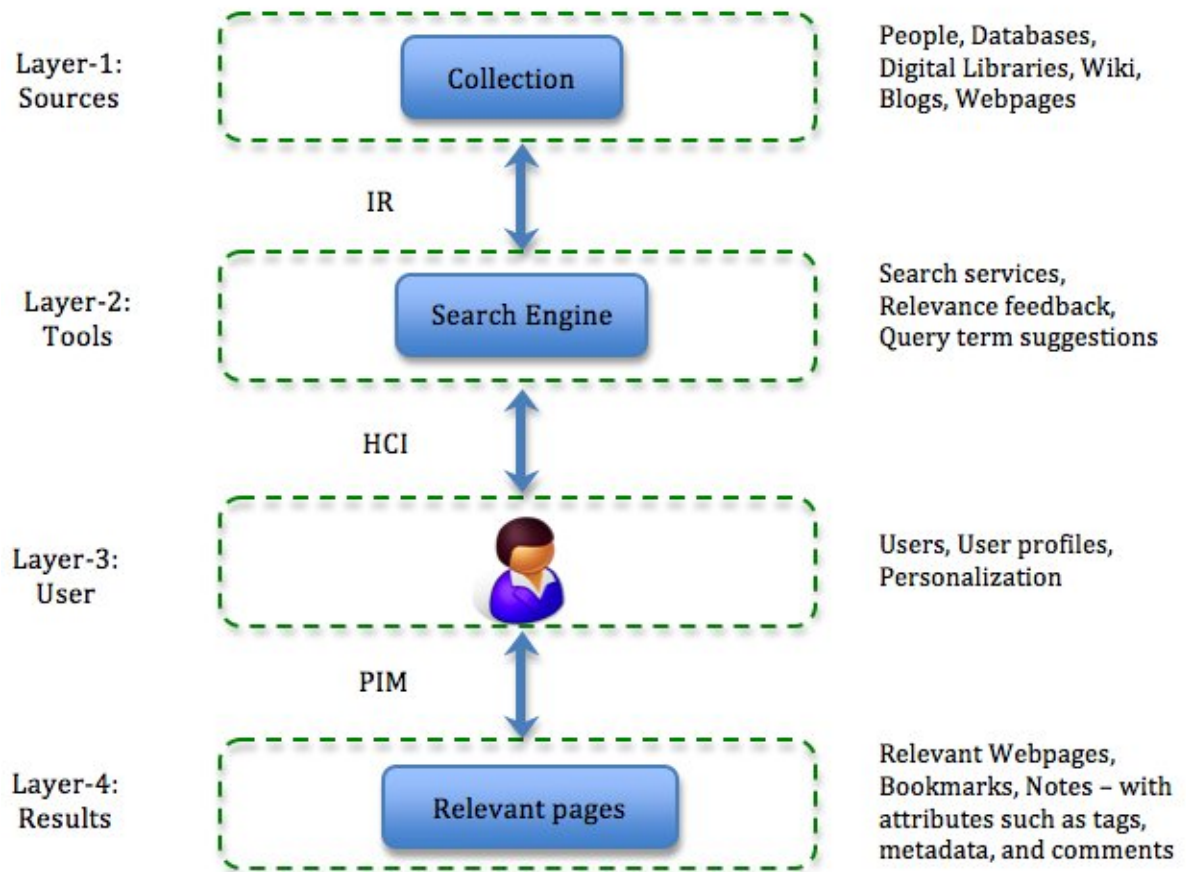


Figure 2.8: Four layer model of information seeking centered around information access and organization.

2003). In addition, since this layer also acts as a mediating layer between information sources and users, it includes a variety of user interfaces, starting from results as rank-lists to touch panels with mechanisms to visualize results. We can see that a large amount of research in IR is focused on the link between layer-1 and layer-2; that is, developing tools and services appropriate for retrieving information of various forms.

Layer-3: User

This layer consists of a user, who uses the tools in layer-2 to access the information in layer-1 and accumulate the knowledge in layer-4. We can see that the focus of Human-Computer Interaction (HCI) research has been on the link between layer-2 and layer-3; that is, presenting the information and the information access tools in effective ways to the user. This layer-3 also includes elements relating to a user, such as user profiles, which can be used for personalization (Teevan, Dumais, and Horvitz, 2005).

Layer-4: Results

The user of layer-3 accumulates the information relevant to him in layer-4. In the most basic sense, this could be a set of webpages that the user found relevant from his searches on the web. Extending this further, we can have bookmarks, notes, and other kinds of results, sometimes stored with attributes such as tags, metadata, and comments. At a more conceptual level, this layer consists of the knowledge that the user gained by his information seeking process. The focus of research in personal information management (PIM) (Dumais, Cutrell, Cadiz, Jancke, Sarin, and Robbins, 2003) has been on the link between layers 3 and 4, addressing the issues of information storage and organization by users.

Let us now see how each of these layers can be extended by allowing various entities at those layers collaborate, and how other fields play out in this extension (see Figure

2.9).

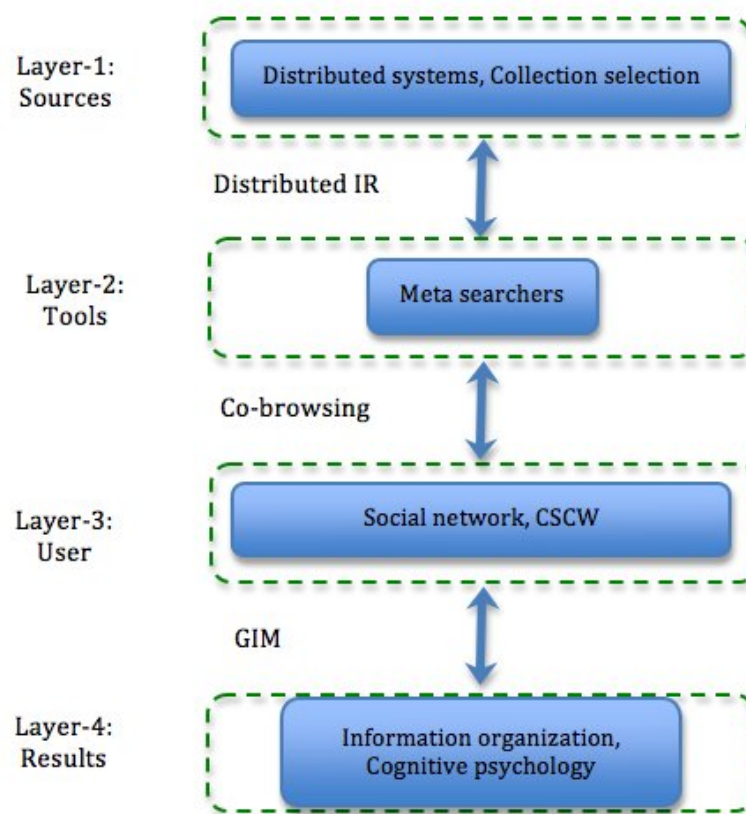


Figure 2.9: Extension of the four layer model of information access and organization.

Layer-1. Combining various information sources is covered under shared resources and distributed systems. From the IR perspective, collection selection also becomes an important issue to address here.

Layer-2. Combining two or more search engines for search process creates meta-search process (Aslam and Montague, 2001). Clusty⁴ and Dogpile⁵ are good examples of such

⁴<http://www.clusty.com>

⁵<http://www.dogpile.com>

meta-search engines.

Layer-3. Putting more than two users together can create a workgroup. Users in a workgroup who access and explore some information together is studied as co-browsing or social navigation (Gerosa, Giordani, Ronchetti, Soller, and Stevens, 2004). Extending the workgroup further with connections among users that exhibit some social characteristics, creates a social network. When multiple user profiles are considered as one entity to optimize the services for that group, personalization is transformed into *groupization* (Morris, Teevan, and Bush, 2008).

Layer-4. Combining relevant information of various sorts falls under information organization and cognitive psychology. Customizing information for a group of users is the topic of research named as Group Information Management (GIM) (Erickson, 2006).

Based on various combinations at individual layers and various kinds of interconnections among these layers, we can imagine a large array of possible scenarios. Discussion on these scenarios is beyond the scope of this work. Instead, possible combinations at the user layer will be explored. There are several ways and reasons the users can connect with each other in an information seeking process. The kind of collaboration considered here is *intentional*, and *interactive* among users with the *same information goal*. In order to create and study such an environment, the general model of information access and organization given in Figure 2.8 will be used and stripped down to suit the needs.

To be specific, the situations in which a group of users collaborate using traditional or collaborative tools to achieve personal or common information goals will be investigated. A model with such a configuration is given in Figure 2.10. This is obtained

by extending the original model of information access and organization for two users.⁶ These users can access and organize information individually, or decide to collaborate with each other. In the case of collaboration, they will have a way to communicate with one another. They may have a common or shared interface. They may also have a shared space where they can store and organize their results. In other words, collaboration between these two users can occur at various levels: (1) while formulating an information request, (2) while obtaining the results, and (3) while organizing and using the results.

The model described above should help in clarifying the presented approach to CIS with the kind of information seeking process, centered around information access and organization. Later while discussing specific approaches to address various issues in CIS, we can revisit this model and see how those other works fit in here.

Looking back to the three phases of collaboration proposed by Gray (1989, p.57-94), and the eight subprocesses of information seeking proposed by Marchionini (1995), we can identify corresponding phases, adding a phase for aggregation, for a collaborative information seeking endeavor as given below.

1. *Pre-negotiation or problem-setting phase*

- (i) Recognize and accept an information problem: this stage presumably brings a set of people together for a possible collaboration as they identify their common goals.
- (ii) Define and understand the problem: the group can discuss the problem and if required, negotiate for a common understanding of the problem. (Figure 2.10,

⁶Considering only two users is merely for the simplicity; theoretically, the model can incorporate more than two users. However, in practice, beyond a certain number of users we may start seeing a point of diminishing return. Worse yet, having too many users for a collaborative project may decrease the productivity of the group. Morris (2008) found from her survey of knowledge workers that the typical size of a collaborative group was 4 to 6 participants.

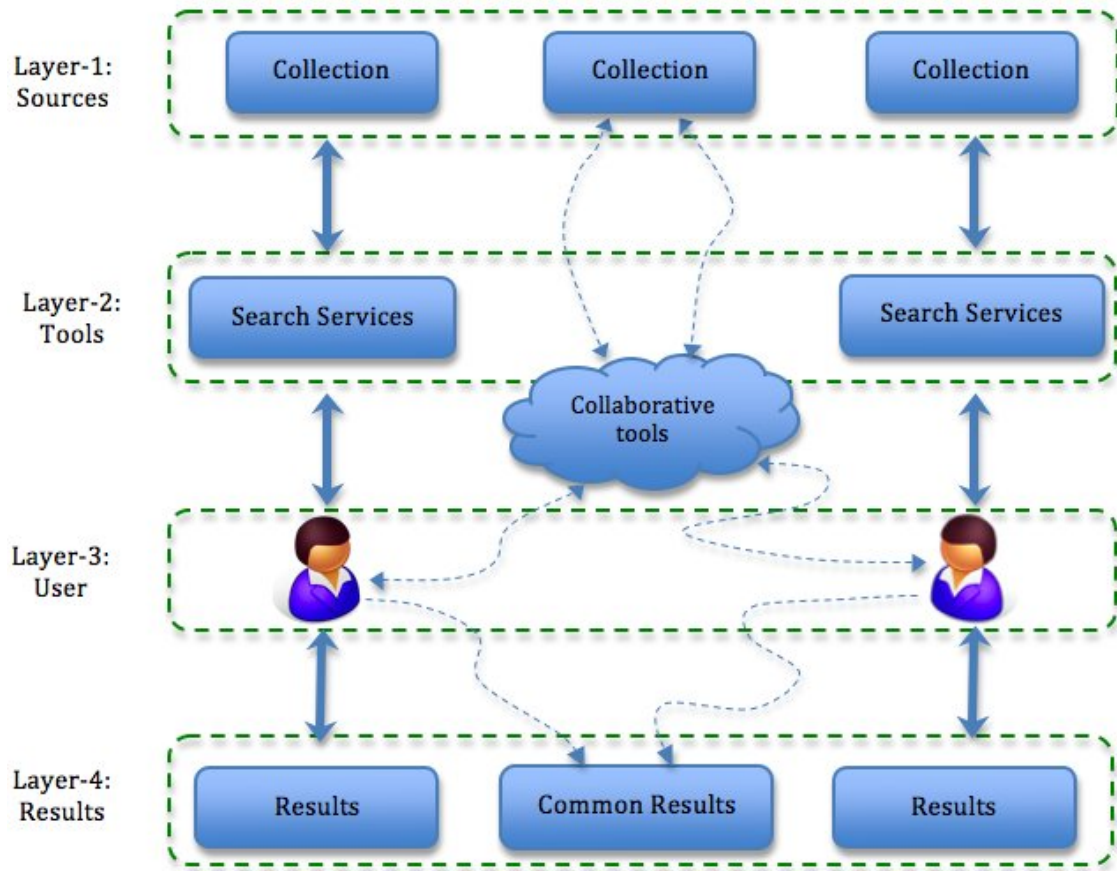


Figure 2.10: A model for Collaborative Information Seeking (CIS) extended from the model in Figure 2.8.

Layer-3)

(iii) Choose a search system: the group can choose source(s) to use and if possible or desired, divide them up among themselves. (Figure 2.10, Layer-1)

2. *Direction-setting phase*

This is an important phase that is more specific to the collaboration. In this phase, the participants need to identify the interests which brought them together, determine how they differ from the interests of the others, set directions, and establish shared goals. (Figure 2.10, Layer-3)

3. *Implementation phase*

- (iv) Formulate a query: given the participants are allowed to work independently, each one formulates his/her search request. At this time, it is beneficial to be aware of the search requests formulated by other participants in the group (Morris and Horvitz, 2007). (Figure 2.10, Layer-3)
- (v) Execute search: everyone executes his/her request on the source(s) being used. (Figure 2.10, Layer-2)
- (vi) Examine results: participants can now examine their own results, and if required, results of the other participants. (Figure 2.10, Layer-4)
- (vii) Extract information: required information is extracted. (Figure 2.10, Layer-4)
- (viii) Reflect/iterate/stop: participants can repeat steps (iv)-(vii) or stop. They may also choose to annotate the results.
- (ix) Aggregate: this phase is more specific to the collaboration, in which the acquired information, processes, and *knowledge* are aggregated to create a shared information and understanding. (Figure 2.10, Layer-4)

2.3 Information filtering

Information filtering refers to a variety of processes involving the delivery of information to people who need it (Belkin and Croft, 1992). In other words, information filtering is a process through which information is derived based on relevance to a user as well as his preferences or past behavior. The manifestation of relevance, preferences, or the past behavior can be limited to the given user or can be extended to map to the same attributes about other users.

To clarify this point, let us plot various scenarios of information seeking. Figure 2.11 shows a typical information seeking scene for a single user. The need for information is expressed and executed, and the found results are returned to the user. The user then

evaluates the results and keeps the ones that are relevant to him.

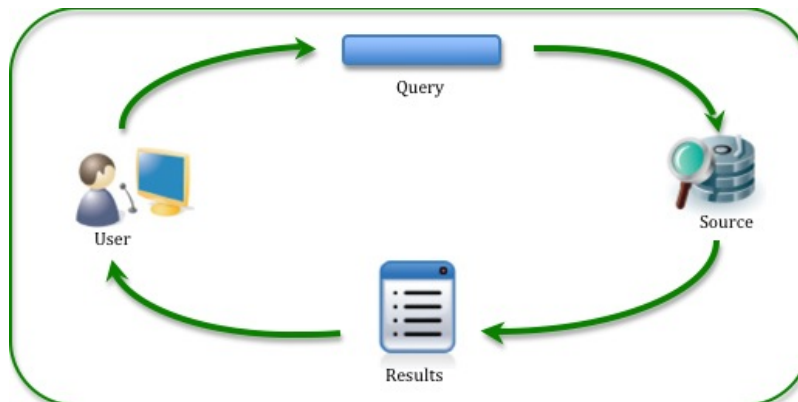


Figure 2.11: A typical scenario of information seeking in an IR environment.

Now, if we had a “smart” system, it will monitor this user’s behavior over time and use it in new information seeking processes (Figure 2.12). The behavior refers to the kind of queries that the user submits, the results that he views, and the information that he saves. In other words, this smart system learns the user’s information seeking model and uses it to aid in future information seeking processes. An example is online movie renting services such as Blockbuster and Netflix. Based on the kind of movies a user has rented in the past as well as the ratings that he assigned, the system recommends new movies to him. In recommender systems literature such an approach is referred to as the content-based recommendations (Adomavicius and Tuzhilin, 2005). In such systems, the utility $u(c, s)$ of item s for user c is estimated based on the utilities $u(c, s_i)$ assigned by user c to items $s_i \in S$ that are “similar” to item s . In our example of online movie renting, in order to recommend movies to user c , the system tries to understand the commonalities among the movies user c has rated highly in the past (specific actors, directors, genres, subject matter, etc.).

In conjunction with the system learning the user behavior, the user himself can also provide his preferences by setting up his profile. The system can then filter the information based on the user profile.

Now let us extend the above scenario to incorporate multiple users. Such a scenario is depicted in Figure 2.13. As we can see, now the system uses the information seeking behavior of other users to aid the given user in his process of information seeking. Typically, these mappings are done based on like-interests among the users. For instance, if users A and B both liked x , and if A liked y , there are good chances that B will also like y . Thus, the system will recommend y to B . In the literature, such kind of recommendation is called collaborative recommendation (Adomavicius and Tuzhilin, 2005).

It is also possible to combine the two kinds of approaches described above combined to create a hybrid approach.

As we saw earlier, collaboration among users based on common information need and asynchronous communication flow results in collaborative filtering. In other words, collaborative filtering refers to a process in which a user benefits from other users' past actions on the same/similar information seeking tasks. In practice, this concept is realized as recommender systems. Examples of such applications include Amazon.com (Linden, Smith, and York, 2003), movies recommendations by MovieLens (Miller, Albert, Lam, Konstan, and Riedl, 2003), and news filtering by VERSIFI Technologies (Billsus, Brunk, Evans, Gladish, and Pazzani, 2002).

Often the systems that are built for supporting collaborative filtering are promoted as CIS systems. A typical characteristic of such systems is a way to combine information requests and/or results in some way. However, due to asynchronous and unidirectional interaction nature of these systems, we cannot call them fulfilling the requirements for CIS that is of interest in here. An example of this can be found in (Klink, 2001), where a method for improving the original query by an automatic reformulation method is proposed. This method uses the term-concept correspondence learned from the documents given by the feedback of the actual or of the other users. Here, there is no direct interaction among the users of the system to carry out a common goal. Thus,

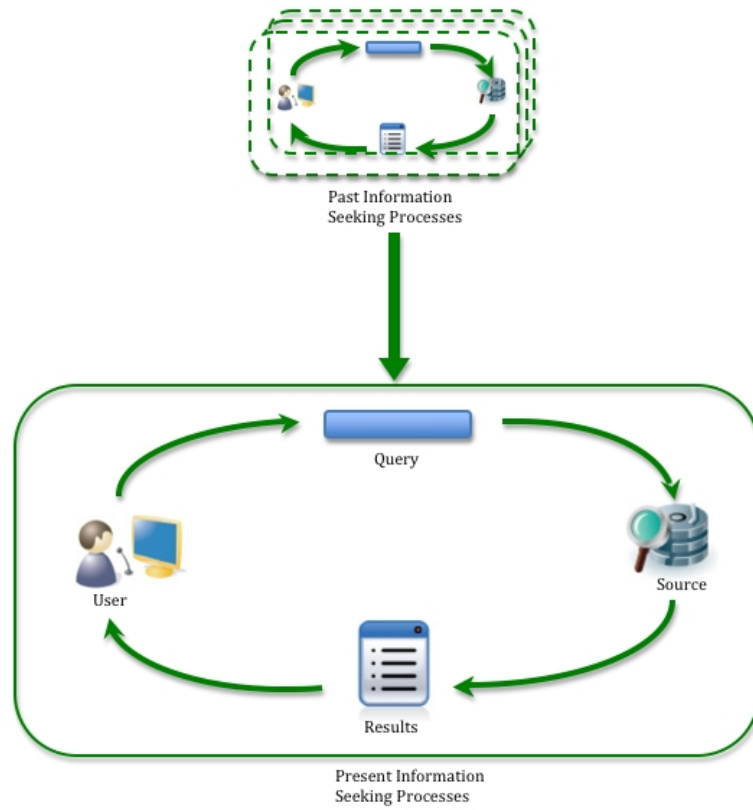


Figure 2.12: Content-based information filtering.

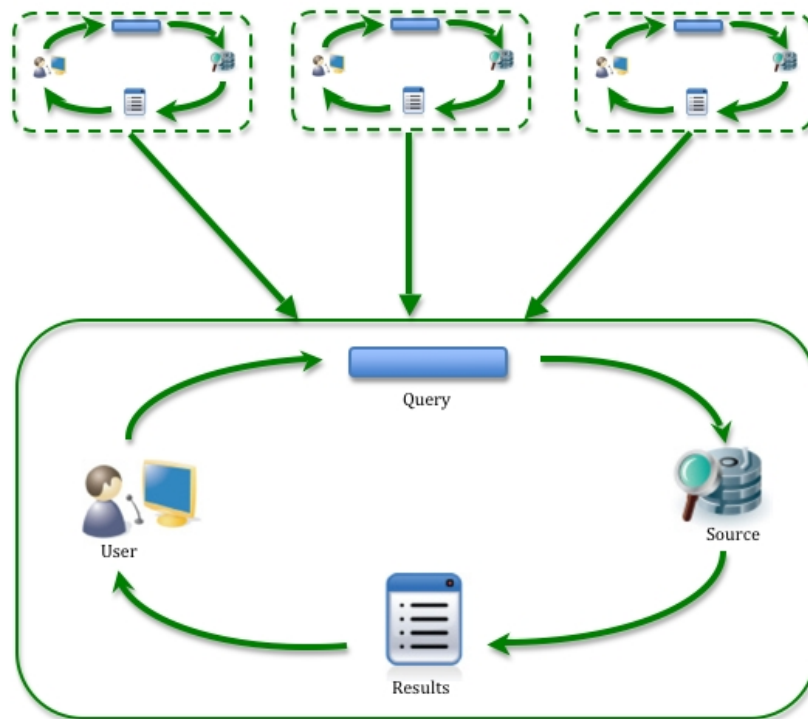


Figure 2.13: Collaborative information filtering.

even though the users are benefitting from their own or other users' past behavior (content-based or collaborative filtering), they are not collaborating in a strict sense.

Furnas (1985) demonstrated the power of community knowledge and collaborative filtering with his adaptive indexing scheme, which helps re-weight indexing terms by the past usage of more experienced users. Along the line of reformulating the queries, Hust, Klink, Junker, and Dengel (2002) showed how to use previously learned queries and their relevant documents for improving overall retrieval quality of an issued query. They do so by expanding the new query by extracting terms from documents which have been judged as relevant to previously learned queries. This approach is further formalized in (Hust, Junker, and Dengel, 2004), and (Hust, 2005). Once again, we see that objects and actions from the past are used to improve retrieval effectiveness, but the “interactions” among the users were asynchronous and unidirectional.

Similarly, Romano Jr., Roussinov, Nunamaker, and Chen (1999) presented *Collaborative Information Retrieval Environment (CIRE)*, a system architecture constructed using the user experiences with IR and Group Support Systems (GSS). Their prototype system employed a client-server architecture, where the server was responsible for connecting the client requests to the AltaVista search engine and recording all the interactions as well as annotations by the participants. The executed queries, stored search results, annotations, and relevance judgments are shared among the group members to facilitate collaborative IR. This is similar to Smyth's I-Spy system,⁷ in which the participants of a workgroup can benefit from others' searches on the same/similar topics. Once again, such systems or environments facilitate collaborative filtering rather than an active and interactive collaboration among the information seeking participants.

How information filtering relates to CIS

⁷This has been transposed to HeyStaks (<http://www.heystaks.com/>).

It is important to note here that in many of these applications, a user receiving the recommendations may not know the other users in the network personally. Thus, a user is not *intentionally* and *interactively engaged* in a true collaboration with other users; he is merely getting filtered content based on other users' actions on the similar information. There have been some applications that go beyond such a constraint and exploit more tightly connected social networks of a user instead of the entire network to filter and recommend information. For instance, Kautz, Selman, and Shah (1997) presented *ReferralWeb*, which was based on providing recommendations via chains of named entities instead of anonymous users in the network.

In addition to this difference, the goal of collaborative filtering systems is to use the opinions of a community of users to help individuals identify content of interest with ease from a potentially overwhelming set of choices (Resnick and Varian, 1997). Looking carefully, we can see that the flow of information and the direction of filtering at a given time is only one way (shown in Figure 2.13). Collaboratively seeking information, on the other hand, involves both the agents actively engaging in an information sharing situation; thus making the flow of information and filtering in both the directions at any given moment.

Another characteristic of a typical information filtering system is the asynchronous nature of user interaction. A user of an information filtering system is provided with the filtered information based on the actions that other users took in the past. In contrast, the focus of the work reported here is on active user interaction among the users of a system who are working in the *same time frame*.⁸

⁸It is not implied here that the interactions should be strictly synchronous; they simply need to be in the *same time frame* (see Appendix A for the definition), letting the users work synchronously or asynchronously as needed.

2.4 User and system interaction

Interactions between users and systems in an information seeking environment are crucial and sometimes inevitable for effective information retrieval. To put the interactions in a single-user or collaborative environment in perspective, several possible scenarios are presented below.

1. *Person to person.* This is a simple case of person to person communication. Almost any dialogue between two people can be considered as an example of such an interaction.
2. *System to system.* This is a simple case of two systems interacting with each other without any human intervention. Consider the following communication between a client and a Network Time Server (NTS).

Client: Hello, I need to know the time.

Server: OK. What's your time-zone?

Client: It's EST.

Server: Then it's 10:43:56 where you are. And it took 0.16 seconds for your message to arrive, so don't forget to account for that time-lag.

Client: Got it. Thanks.

Of course, the actual communication between the client and server is different than this, but the essence of this description is that two autonomous entities are interacting without any feedback from a human.

3. *Person to system.* We encounter this scenario quite often in day-to-day life when we have an interactive session with a system. Consider using *Lygos*,⁹ an interactive search engine. You type in a query and it reports that you have a typo

⁹*Lygos* has nothing to do with Live, Yahoo!, Google, Lycos, or any such search engines!

in your query. You correct that and try your search again. You are not satisfied with your search results, and so you try a different query. Thus, the process of accomplishing some task may take several cycles of interactions.

4. *Person to person mediated by a system.* This is a scenario when two users have a system between them, mediating the interactions. IM or online chatting is a good example of such an interaction. This is also the scenario that is studied very frequently in IR, more specifically for relevance feedback.
5. *System to system mediated by a person.* It is not clear if the scenario where a user is mediating two systems is practically useful or not, while it is certainly possible.
6. *Group of people.* This scenario covers our traditional social and economical bonds where a group of people interact with each other as they belong to a social, economical, political, or ethnic class.
7. *Group of systems.* Similar to the previous scenario, where a group of people created a network, a group of systems can be connected to form a computer network. We see such networks everywhere around us - as the Internet, for Cloud Computing (Wikipedia, 2008), or to facilitate parallel and distributed processing and storage.
8. *Group of people mediated by a system.* Combining scenarios 3 and 6, we can have a situation where a group of people are interacting with each other mediated by a system. Most of the online social networking sites can fall under this scenario.

Since the focus here is on studying computer-supported collaboration, the process of CIS, by the given definition, involves at least two users and one intermediary system. As shown earlier in Figure 2.10, the users of a CIS environment could communicate with each other through the system or interface provided to them, but we can also

imagine having interactions that do not require such an intermediary if the users are co-located. For instance, Root (1988) showed how exchange of ideas around the coffee machine is a popular way of collaborating with other people, often without using any system. This suggests that while designing a CIS system, we should be open to the possibility of different forms of interactions: human to human, human to system, and human to human mediated by a system.

Consider the user interactions with IR systems. Saracevic (1996) provided a good review of various models of such interactions. To be specific, he identified three models.

1. *Traditional model.* This model, as he said, “...represents IR as a two prong set (system and user) of elements and processes converging on comparison or matching...”
2. *Ingwersen’s cognitive model.* This model (Ingwersen, 1996) “...concentrates on identifying processes of cognition which may occur in all the information processing elements involved.”
3. *Belkin’s episode model.* This model (Belkin, Cool, Stein, and Thiel, 1995) “..considers user interaction with an IR system as a sequence of differing interactions in an episode of information seeking...”

Saracevic himself then provided a model, which he called the *stratified interaction model* (Figure 2.14). This model was developed within an overall framework of an ‘acquisition-cognition-application’ model of information use. The levels or strata posited by Saracevic are simplified to three: *surface*, or the level of interaction between the user and the system interface; *cognition*, or the level of interaction with the texts or their representation; and the *situation*, or the context that provides the initial problem at hand.

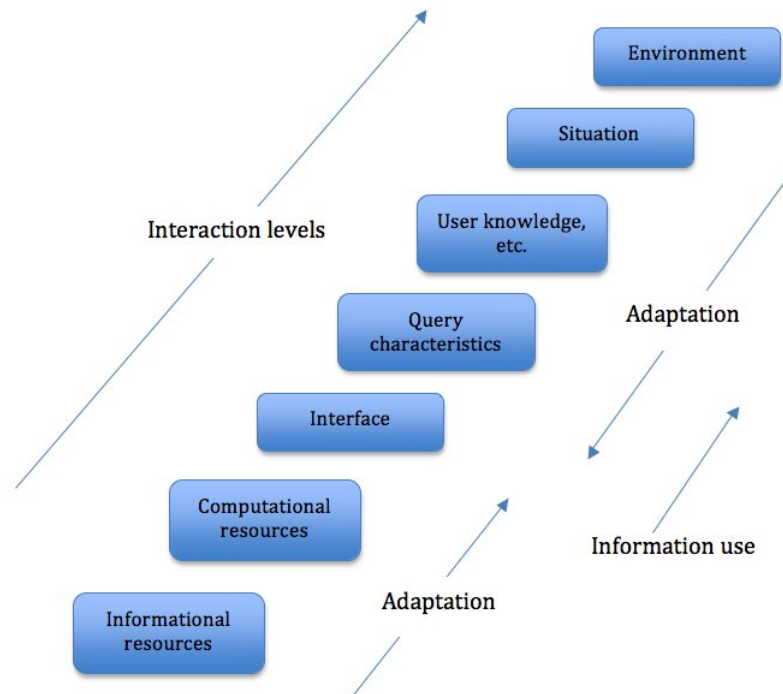


Figure 2.14: Saracevic's model of stratified interaction.

With such a stratified model with multiple interaction and processing strata for the user and for the system, Saracevic (1997) suggested that the user and the system have equal constraints on the search. The system side of this model was later extended by Spink, Greisdorf, and Bateman (1998), who added a graduated relevance dimension, and Bates (2002), who identified additional levels that interact and affect each other. The key idea of these stratified models of interaction is that even if all other components but one is executed effectively on either user or system side, the component that was done poorly can hinder the value of the entire system. For instance, a user's interpretation of results may seriously affect their success in achieving their goals with the system. This allows us to evaluate every layer of the model to check for the bottlenecks in the performance. However, stratified models themselves do not provide a way to evaluate the layers. This and some of the other shortcomings of the stratified models were addressed by episodic models proposed by Belkin et al. (1995). These kind of

models define the flow in scenarios of human-system interactions. The flow definitions are called “scripts”, which define the typical steps of interaction between a user and a system. To do this, Belkin et al. proposed to use four binary dimensions - *Method*, *Goal*, *Mode*, and *Resources*. Possible combinations of these four dimensions produce what they called Information Seeking Strategies (ISS), presented in Table 2.2. They calculate separate scripts for each of these 16 ISS conditions, which allow for switching between them.

Table 2.2: Information seeking strategies (Belkin et al., 1995).

ISS	Method	Goal	Mode	Resources
1	Scan	Learn	Recognize	Information
2	Scan	Learn	Recognize	Meta-Information
3	Scan	Learn	Specify	Information
4	Scan	Learn	Specify	Meta-Information
5	Scan	Select	Recognize	Information
6	Scan	Select	Recognize	Meta-Information
7	Scan	Select	Specify	Information
8	Scan	Select	Specify	Meta-Information
9	Search	Learn	Recognize	Information
10	Search	Learn	Recognize	Meta-Information
11	Search	Learn	Specify	Information
12	Search	Learn	Specify	Meta-Information
13	Search	Select	Recognize	Information
14	Search	Select	Recognize	Meta-Information
15	Search	Select	Specify	Information
16	Search	Select	Specify	Meta-Information

Let us see how these ISS conditions inform user interactions in various information seeking situations. A typical information seeking situation is a user searching on the web with a search engine. This interaction can be fit in ISS15, where the user is searching (Method) to select (Goal) by specifying (Mode) attributes of a specific information object (Resource). For collaborative information seeking, we can imagine the users going through various forms of such interactions, the knowledge of which could provide

us with enough information to evaluate the quality of the whole information seeking experience.

How user and system interaction relates to CIS

The models that are listed above give us plethora of information for observing and evaluating various forms of user and system interactions. However, they do not tell us enough about the degree to which a user or a system is involved. While it is also assumed that the interactions are intentional or explicitly specified, we can imagine a gradation of involvement for both the user and the system in an interactive environment. Such understanding is important to extend interactions among users and systems to collaboration. To address these issues, two different ways of looking at the interplay of user and system in collaboration are presented below.

2.4.1 Degree of involvement

Figure 2.15 presents collaborative systems in the context of the degree of involvement that a user or system component has. The four quadrants, starting top-right and going in the counter-clockwise direction are explained below.

- The first quadrant shows the applications where a high degree of interactions among users and systems produce collaboration. In the case of email or a video conference, there is an end-to-end collaboration between users and it is facilitated by certain protocols and communications between systems of those ends.
- The second quadrant shows the situations in which the users have a high degree of interaction to facilitate the collaboration, but the systems do not play much role. For instance, an exchange of ideas around the coffee machine is a popular way of collaborating with people, often without using any system (Root, 1988).

- The third quadrant represents most of the IR systems that are designed for a single user.
- The fourth quadrant shows the applications where there is a higher degree of interactions among systems than among users. For instance, meta-search engines, such as Clusty, allow a single user to search different systems and combine the results. In the far right corner of this quadrant, we can also see applications such as Cloud Computing (Wikipedia, 2008) and distributed IR (Callan, 2002).

2.4.2 Explicitness of collaboration

Figure 2.16 presents various collaborative systems in the light of how explicit or intentional a user or system component is. The four quadrants, starting top-right and going in the counter-clockwise direction are explained below.

- The first quadrant shows applications where both the system and the user explicitly collaborate. For instance, in case of a patron walking up to the reference desk in a library for an information need, there is an explicit collaboration between users and some degree of explicit collaboration with the system (e.g., using OPAC). In case of phone directory, a user agrees to put his information public; he and others receive the directory that has everybody's public information.
- The second quadrant has applications where there is more explicitness for collaboration from the user side and less from the system side. For instance, the *Ariadne* system (Twidale and Nichols, 1996) allows two users to explicitly collaborate while handling a lot of aspects of collaboration implicitly. Another example in this quadrant is a social bookmarking service such as del.icio.us, where the users store bookmarks, not only for their own reference, but also to share with others. The system handles this aspect of collaboration internally.

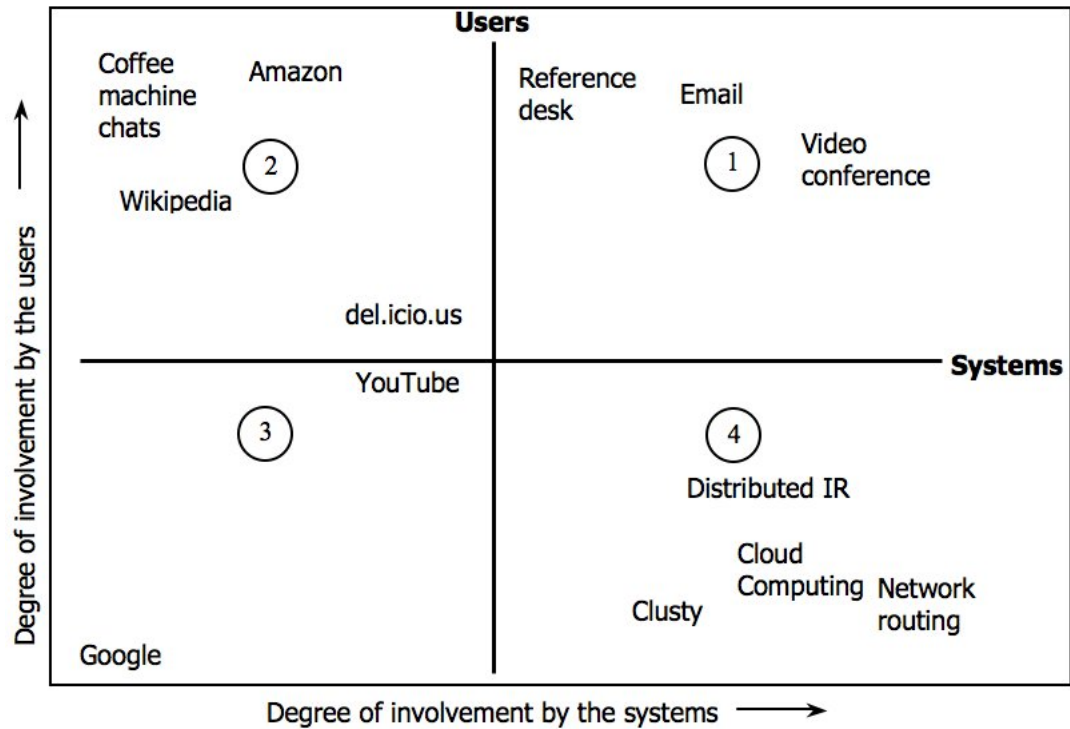


Figure 2.15: Looking at collaboration with the amount of user and system involvement.

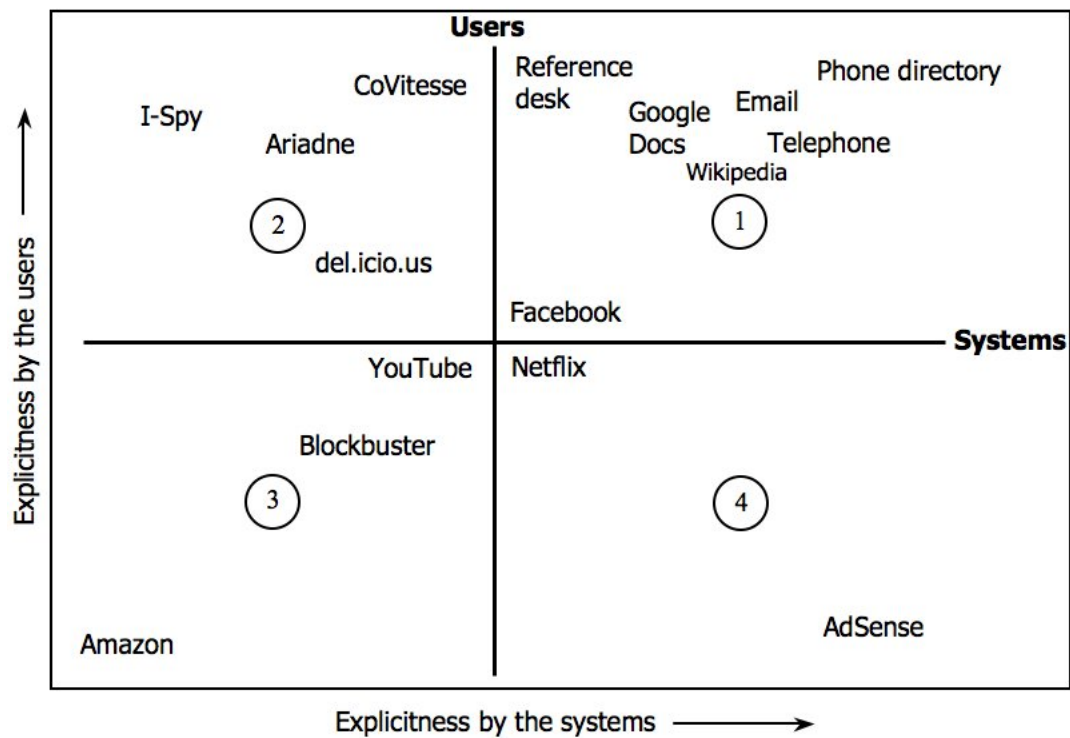


Figure 2.16: Looking at collaboration with the amount of explicitness or intention from user and system.

- The third quadrant shows applications where the users’ collaboration as well as system’s implementation of this collaboration is more implicit. For instance, users of Amazon.com receive certain information that is influenced by other users, but not formally defined or intended.
- The fourth quadrant shows applications where implicit collaboration among the users is done explicitly by the systems.¹⁰ For instance, users of Google implicitly agree to receive ads based on their searching; Google puts targeted ads (AdSense¹¹) on search pages.

The work presented in this dissertation falls somewhere in the top left corner (quadrant two) of Figure 2.15 as well as Figure 2.16, where the participants are controlling the collaboration with intention and interactions among themselves, with appropriate support implicitly provided by the system. In fact, it is implied that a “true” collaboration occurs among humans, and that there has to be intentionality for doing so. For the work that is reported here, it is useful to think of collaboration as *only* those activities where a group of people have explicitly agreed to work together. This immediately discards any of the collaborative filtering systems that we saw earlier as they do not support “true” collaboration. Thus, a need to differentiate between implicit and explicit collaboration is eliminated, inferring that a “true” collaboration is among humans, and it is intentional or explicit, and interactive.

2.5 Social networking

Social networking services using software that allows a group of people to build their online social network is a relatively recent concept. However, the notion of social

¹⁰See Appendix A for the definitions of implicit and explicit collaborations.

¹¹<https://www.google.com/adsense/>

networks is not new.

A social network is a structure made of nodes (of people, organizations, or other entities) and connections among these nodes. The connections could be based on several factors including friendship, kinship, trade, and web links. People have studied such social structures for centuries. However, researchers in social network analysis typically recognize Jacob L. Moreno's work on sociometry (Moreno, 1934) as pioneering in this field. Freeman (1996), though, points to some earlier works that also deserve credit for being original in the field of social network analysis.

Many of these earlier works had one core idea in common - *homophily*. The theory of homophily, defined by Lazarsfeld and Merton (1954), says that most human communication will occur between a source and a receiver who are alike, that is, homophilous, and have a common frame of reference. The idea of *birds of a feather flock together* itself is not new. More than two millennia back, Plato (427 BC - 347 BC) noted that "*similarity begets friendship*". His pupil, Aristotle (384 BC - 322 BCa), also observed that people "*love those who are like themselves*." Referring to more contemporary works, Deschields and Kara (2000, p.315) recognized that "*The earliest studies of homophily concentrated on small social groups, in which an ethnographic observer could easily ascertain all of the ties between members (whether those ties were behavioral, like sitting together at a cafeteria table, or reported, as when an informant tells about his or her close friends). Therefore, our first systematic evidence of homophily in informal network ties came from school children, college students and small urban neighborhoods. (p317)*"

There are many studies that support such theories. Almack (1922) asked children of age 4 to 7 in a California school to answer a series of questions about those with whom they would like to work and those with whom they would like to play. Almack tabulated chooser-chosen pairs and then ran correlations between the I.Q.'s of

pairs. His examination of these correlations confirmed the hypothesis that choices are *homophilous*.

Similar to Almack, Wellman (1926) focused on homophilous choices among pairs of individuals. The difference came from her approach to studying such social behavior. Unlike Almack, who depended on reports of choices, Wellman recorded those pairs of individuals whom she had observed as being together frequently. Along the same line, Bott (1928) observed children at the preschool attached to the University of Toronto and found that five forms of behavior recurred regularly among the children: (1) talked to another, (2) interfered with another, (3) watched another, (4) imitated another, and (5) cooperated with another. These studies are highly relevant to studying CIS environments. For instance, while the participants are engaged in a CIS environment, we can measure how often (1) these participants communicate with one another, (2) take some action based on their interactions, (3) agree with one another, (4) conflicted with one another, and (5) distributed responsibilities or combine individual contributions.

An important issue in studying social networks is how various forms of interactions take place among the participants. Hubbard (1929) pioneered in the study of techniques for observing interaction. She systematically examined inter-observer reliability in recording patterns of who interacted with whom among a set of preschool children. Hagman (1933), following this track, observed interaction frequencies among children - who played with whom - repeatedly over a school term. Then she interviewed her subjects and asked them to recall their playmates of that day, those with whom they played the day before, and those with whom they played at the beginning of the school term. All of these studies relating to school children once again confirmed the hypothesis of homophily in different forms. We find this line of research being followed in recent times, for instance, analyzing children's friendship relations, Newcomb and Bagwell (1995) found that friendships, compared with nonfriend relations, are characterized by

more intense social activity, more frequent conflict resolution, and more effective task performance. Such works are helpful in understanding how and why people collaborate with each other. It seems from these studies that people are more likely to collaborate with others who are alike (homophilous). However, this idea of homophily does not imply that people come with the same kind of backgrounds and skills. In fact, it is the diversity that helps make collaboration more meaningful (Surowiecki, 2004). In other words, while people tend to collaborate with those who are like them (probably with some common goals), they benefit more when there is diversity of opinions and skills in the group.¹²

A parallel line of research spawned around the idea that peer groups were an important source of influence on people's behavior, especially among young people. This influence could be positive or negative and the study of such influences were done with cross-sectional association between a set of characteristics of an individual and the corresponding characteristics of that individual's friends (Berelson, Lazarsfeld, and McPhee, 1954). This is an important finding for the kind of collaboration that this dissertation addresses, namely explicit, interactive, and mutually beneficial. In contrast to collaborative filtering or implicit collaboration, the users of "true" collaboration are actively involved influencing and benefitting from one another in the group.

Understanding how social networks formulate and function is a very active line of research and the area of social network analysis is enriched with works from sociologists, economists, anthropologists, information scientists, and many other professionals (Krebs, 2006). Instead of dwelling on further depths of this field, a discussion on social

¹²At the same time, it is acknowledged that often workgroups are forced, and not naturally created. Once again, it is important to note that the focus here is on studying collaborations that are intentional and explicitly defined.

network services follows.¹³ Due to their ever-growing use and success, such social networks help us design CIS systems with a better understanding of social interactions. For instance, a survey by Evans and Chi on Amazon’s Mechanical Turk service (Evans and Chi, 2008), suggested that social interactions play an important role throughout the search process. They used their survey data to propose a social model of user activities before, during, and after search.

A social network service is used to build online social networks using software. The typical goal of creating such networks is to connect a group of people who share interests and/or activities. Early implementations of social networks can be seen in Classmates.com (1995) and SixDegrees.com (1997).¹⁴ In less than a decade’s time, social networking sites have found their functionalities in not only connecting friends, but also in business (e.g., LinkedIn.com), and health-care (Luo, 2007). Some of the social network services are also noted to be useful in legal investigations. For instance, content posted on MySpace helped expose sex predators (France-Presse, 2007).

Millions of users around the world are using a variety of such networks, which are primarily web-based, to interact with each others using chat, messaging, email, video chat, file sharing, blogging, and discussion groups. The popularity of various social networks in different regions of the world is given below.

- *North America:* MySpace, Facebook (Roxborough and Masters, 2007)
- *South and Central America:* Orkut (Liebel, 2008), Hi5
- *Europe:* Bebo (O’Hear, 2007), MySpace, Skyrock Blog, Facebook, Hi5
- *Asia and the Pacific Islands:* Friendster, Orkut (Maderazo, 2007)

¹³For a more comprehensive treatment of social network analysis, the reader is referred to Wasserman and Faust (1994) and d. m. boyd and Ellison (2007).

¹⁴This site was eventually shut down due to financial problems. The owner declared it was “simply ahead of its time.”

What this shows us is that while there may be cultural differences in the way people work together in different systems and parts of the world, they all have access to some form of social network service that connects them with their peers, friends, and families. Establishment of trust with other nodes on the network is one of the first steps to a successful collaboration. In addition to this, studies of such online social networks also involve issues such as privacy, which is an important factor to consider while designing CIS systems. Just as in an online social network environment, people are willing to share certain information, but do care about their privacy, thus we need to make sure that the users of a CIS system have control over what information about them gets shared with the other participants in the group.

How social networking relates to CIS

Collaboration is a complex process. While collaboration can happen informally and in an impromptu fashion, often it is a process that evolves over a period of time. During this time the interactions and ties among the participants may also change due to the demands of the collaborative project and/or simply because of social aspects of these interactions and ties.

As depicted in the set-based model for collaboration earlier, a social interaction starts with as little as simple communication. It is possible that an interaction can end at that level, or can also move forward to other levels evolving all the way to collaboration. It is important to note that the proposed model is not intended to show all possible paths starting with communication. For instance, a communication can lead to a new friendship without going up the path of collaboration.

Understanding how these interactions take place at the communication level, and evolve into stronger ties among the participants leading them to collaboration is very useful in not only studying the motivations and scenarios of collaboration, but also in

designing effective systems to encourage and support collaboration.

In order to differentiate five layers of the set-based model of collaboration, and put social and collaborative ties in perspective, three factors have been identified: *objective (what)*, *intention (why)*, and *interactions (how)*. In case of communication without further progress toward any other layers, the participants are simply trying to convey some information (objective and intention) using some form of medium (interactions). As we move toward the collaboration layer, these three factors change. At the coordination layer, the participants are trying to organize actions, events, or objects in some fashion (objective) to have a meaningful outcome (intention). In order to do so, they will exchange information with each other, often going back and forth (interactions). An example is organizing a group meeting.

At the collaboration layer, the participants are trying to solve a problem together or working toward a common goal (objective) for mutual benefits (intention). During this process, they will not only exchange information for communicating or coordinating, but also for sharing opinions, reporting findings, and reflecting on various issues (interactions). Following is a layer-by-layer description of how social networks relate to collaborative groupworks.

Communication

This is an important part of a social network. Often the social ties are formed due to communication. For instance, a friend-request on Facebook is a form of communication, which can result in creating a “friends” tie between the requester and the accepter. However, not all communications result in collaboration. In a social network, if the participants are simply communicating (interaction) without a specific goal or strong intention, the participants may not move to the higher level in the set-based model.

Contribution

In a social network, participants often provide contributions to the people they are tied with. For instance, a comment on a friend's photo on Flickr is a form of contribution. However, this may not result in further progress toward collaboration. In this case, there is interaction, and an intention, but there is no common objective between the person who posted the comment and the person who received it.

Coordination

Social networks are often used to coordinate events. For instance, college students frequently use "Events" feature on Facebook to create an event and invite their friends to it. With each event, the participants have the ability to exchange messages (communication/interaction), and post pictures, videos, or comments (contribution). However, the majority of these interactions and contributions are either of organization and logistical nature, or pertaining to conveying some interesting information without pursuing any higher objectives. The participants may still be connected with weaker ties where they are not working tightly together for solving a problem or achieving a goal, other than getting together for an event.

Cooperation

This is an act of following another participant's rules to get something done. For instance, on WikiAnswers, participants try to address a question they are interested in while following the guidelines of WikiAnswers and the expectations of the asker. Everyone is contributing in a coordinated fashion, but beyond that, they also have to follow the rules set by the system, and the demand by the asker of the question. Finally, it is the asker that benefits the most; the answerers merely cooperated. In

other words, while the level of interactions was higher and there was a clear objective and intention, there was a mismatch in the roles the participants played in this small social network; the asker's intention was to get some answers and satisfy his information need (objective), whereas the answerers' intention was to address that question while following the rules of interactions so that their answer satisfies the asker (objective).

Once again, using the notions of interaction, intentions, and objectives, we can situate the scenarios like the one described above in the set of cooperation.

Collaboration

Social networking with peers can often result into a collaborative project. Several such situations take place in day-to-day life in which during a social interaction with fellow researchers, an idea of collaboration sprung and resulted in a long lasting collaborative project. In such cases, a weaker social tie (colleagues), and unintended interaction (conversation on a topic), resulted in a stronger tie (co-authors), a strong objective (research paper), and intention (working together to address a problem). It is also interesting and relevant to note that such collaborations become possible because of conscious or unconscious awareness of the other people or their actions that one has. For instance, being aware of another researcher's research interests and goals could be a key to initiating collaboration with him. Such awareness may not be required for non-collaborative endeavors.

As we can see, looking at the level of interactions and awareness, intentionality, and the nature of the objective in a social network can help us classify such a scenario as a collaborative one.

Let us see how these factors help us understand social and collaborative ties. A social network is created using general and weaker ties, whereas a collaborative group is formed using more specific and stronger ties. For instance, a social network of a

neighborhood is formed simply based on the fact that certain people live in the houses that are in the same neighborhood. These residents are connected with weaker ties; they often have the same housing-related issues. However, they may not be working together in collaboration with a strong objective or intention.

A pilot and a co-pilot, on the other hand, are working with a much stronger objective (fly the plane) and intention (fly safely and punctually). During the process of flying, they will have tightly integrated interactions. This is an example of collaboration.

One may have hundreds of “friends” on Facebook, but one does not necessarily collaborate with them working on a specific project toward a common and mutually beneficial goal. On the other hand, having such weaker ties through a social network can help in creating and sustaining a collaborative alliance.

In case of interactions relating to searching for information, Evans and Chi (2008), discussed how social interactions could help in searching together. They called this *social search*. Such social ties leading to social search can be extended to stronger ties leading to collaborative search.

Let us talk about such ties in information seeking environments, as transferring such weaker ties to stronger ones to encourage possible collaboration has been used several other places too. For instance, there are co-browsing applications that let visitors of the same webpage be aware of each other, hoping they may want to collaborate as they have the same information need (Donath and Robertson, 1994).

Sometimes the stronger ties are formed not to do collaboration, but for a possible filtering of information. Most of the collaborative filtering systems depend on converting weaker ties (e.g., being the users of the same system and interested in similar objects) to stronger ties (e.g., connecting the users based on their behavior, and having them influenced by each other). For instance, a Netflix user can have social (weaker) ties with his friends on Netflix network, but when Netflix’s collaborative filtering system starts

making recommendations based on one's social network, and when the users in the network start using those recommendations and/or start interacting with their peers based on their similar interests, the weaker ties of social network become stronger and more specific.

In summary, a social network typically exhibits weaker ties among the participants, based on their interactions, intentions, and objectives. A collaborative network, on the other hand, shows stronger ties. A social tie can be useful and converted to a collaborative tie. The reverse can happen too. Often participants without social ties are put in a collaborative project. While working on such a project, the participants may develop a social tie as well. Based on this, it can be seen that one tie (social or collaborative) does not subsume the other; they both can be complementary to each other.

While a social network differs from a collaborative group based on the strength of the ties (which was proposed to be measured by objectives, intentions, and interactions), one can be useful for creating and understanding the other. There have been several works on social networks, and since a social tie can be converted to a collaborative tie, we can learn a lot about collaborative groups by looking at those works done on social network analysis.

As we saw, several of the early works on social network analysis explored the notion of homophily. One of their key findings was that the people tend to create social ties with those who match their interests. While this may not be a surprise, it tells us that in order to have a tie, the participants need to have something in common, and the more the commonalities, the stronger a tie can be. For instance, two users of Twitter may not have any tie at all, but when one discovers the other to be interesting or relevant in some way, he/she can decide to follow that person's Tweets. This creates a stronger tie. Such a tie can eventually be useful for creating possible collaborations. In

our personal experience, we have seen several collaborations happened through blogs and feeds subscriptions (stronger ties).

Another line of research in social network analysis looked at the influence of the peers on a network. Works, such as (Berelson et al., 1954), showed that people are easily and frequently influenced by their peers on the same social network. Such behavior was more predominant in younger generations. Today, online social networking services, such as MySpace and Facebook, make such influences even easier and more frequent.

Considering that a weak form of tie in a social network can be transferred to a stronger tie creating collaboration among the participants, and the participants can be influenced by that weak tie, we can study the motivations (why) and scenarios (what, when) of collaboration by looking at the influences in social networks.

It is also important to note here that one of the interesting factors to study in collaboration is the social aspect of it. Social interactions happening due to the collaboration can be engaging, enriching, and entertaining.

Moving our attention to the technologies that facilitate social or collaborative ties, it should be noted that with the advent of technology involving social networking, people are increasingly becoming familiar and encouraged to shared information about themselves as well as exploring other people's information. Such information exchange is used not only for connecting people or providing recommendations, but also for accomplishing a variety of tasks as we saw earlier. The analysis of social networks, on the other hand, is a well-studied domain for nearly a century and is being adapted to the newly emerged online social networking sites.

What is of interest here is the realization that some of the aspects of CIS, particularly communication, are analyzed extensively in social networking research. In addition to this, research in CIS can also benefit from the understanding developed from social network analysis about how and why people work with each other, the costs and benefits

of such collaborations, and user behavior in these ties or networks.

Collaboration can also be considered as a stronger form of social tie that, according to the definition presented here, involves a group of people working together for a common goal. Often the seeds for such collaboration are planted at the level of social interactions. As Karamuftuoglu (1998) argued, knowledge production, as a part of IR, is fundamentally a collaborative labor, which is facilitated by community interactions. This argument allows us to look at the analysis of social network from a different perspective in which social searching, which is a weaker form of searching together (Evans and Chi, 2008), can lead to collaborative searching, which is a tighter and more specialized form of searching together.

2.6 Collaborative Information Seeking (CIS)

Having reviewed a set of fields that are related to CIS, let us now focus on works that directly address issues of CIS. Before we do so, it is important to point out that the literature is filled with works that use terms such as *collaborative IR* (Fidel, Bruce, Pejtersen, Dumais, Grudin, and Poltrock, 2000a), *social searching* (Evans and Chi, 2008), *concurrent search* (Baecker, 1995), *collaborative exploratory search* (Pickens and Golovchinsky, 2007), *co-browsing* (Gerosa et al., 2004), *collaborative information behavior* (Reddy and Jensen, 2008; Talja and Hansen, 2006), *collaborative information synthesis* (Blake and Pratt, 2006), and *collaborative information seeking* (Hertzum, 2008; Shah, 2008), which are often used interchangeably.

There are several definitions of such related or similar terms in the literature. For instance, Foster (2006, p.329) defined collaborative IR as “*the study of the systems and practices that enable individuals to collaborate during the seeking, searching, and retrieval of information.*” However, there is still a lack of a definition or a terminology that is universally accepted. On top of that, the focus here is not only on retrieving

or browsing in collaboration, but also on performing information seeking in collaboration. As we saw earlier, information seeking goes beyond searching and retrieving information. Nonetheless, an attempt will be made to recognize and classify works by first looking at space-time and user-source-time aspects, and then review them as co-browsing or collaborative IR works.

2.6.1 Space and time aspects of CIS

The classical way of organizing collaborative activities is based on two factors: location and time (Rodden, 1991). Recently, Hansen and Jarvelin (2005) and Golovchinsky, Pickens, and Back (2008) also classified approaches to collaborative IR using these two dimensions of space and time. Figure 2.17 shows various library activities on these two dimensions as proposed by Twidale and Nichols (1996).

As we can see from this figure, the majority of collaborative activities in conventional libraries are co-located and synchronous, whereas collaborative activities relating to digital libraries are more remote and synchronous. Social information filtering, or collaborative filtering, as we saw earlier, is a process benefitting from other users' actions in the past; thus, it falls under asynchronous and mostly remote domain. These days email also serves as a tool for doing asynchronous collaboration among users who are not co-located. Chat or IM (represented as 'internet' in the figure) helps to carry out synchronous and remote collaboration.

The placement of a CIS environment on this figure has implications for its implementation, functionalities, and evaluation. For instance, Adobe Connect¹⁵ facilitates online meetings where the participants can share and discuss information. Such an environment will fall under Synchronous-Remote collaboration in Figure 2.17. Thus, this environment needs to have (1) a way to connect remote participants, (2) a shared

¹⁵<http://www.adobe.com/products/acrobatconnect/>

space for exchanging information, and (3) a communication channel to provide real-time message passing among the participants.

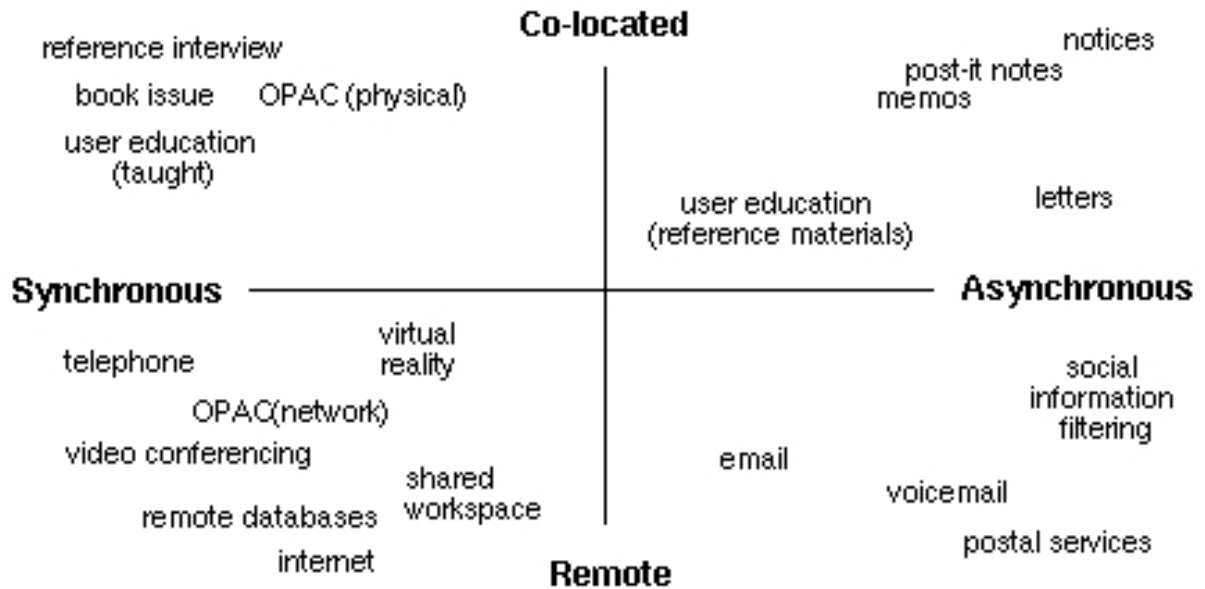


Figure 2.17: Looking at collaboration with space and time dimensions. (Twidale and Nichols, 1996).

Rodden (1991), similarly, presented a classification of CSCW systems using the form of interaction and the geographical nature of cooperative systems (Figure 2.18). Further, Rodden and Blair (1991) presented an important characteristic to all CSCW systems - control. According to the authors, two predominant control mechanisms have emerged within CSCW systems: speech act theory systems, and procedure based systems. These mechanisms are tightly coupled with the kind of control the system can support in a collaborative environment (discussed later).

2.6.2 User-source-time configuration for CIS

Another way of looking at how different systems fit in a broad spectrum of collaboration is to consider these three important aspects: user(s), source(s), and time. These aspects and their corresponding examples are consolidated in Table 2.3, followed by

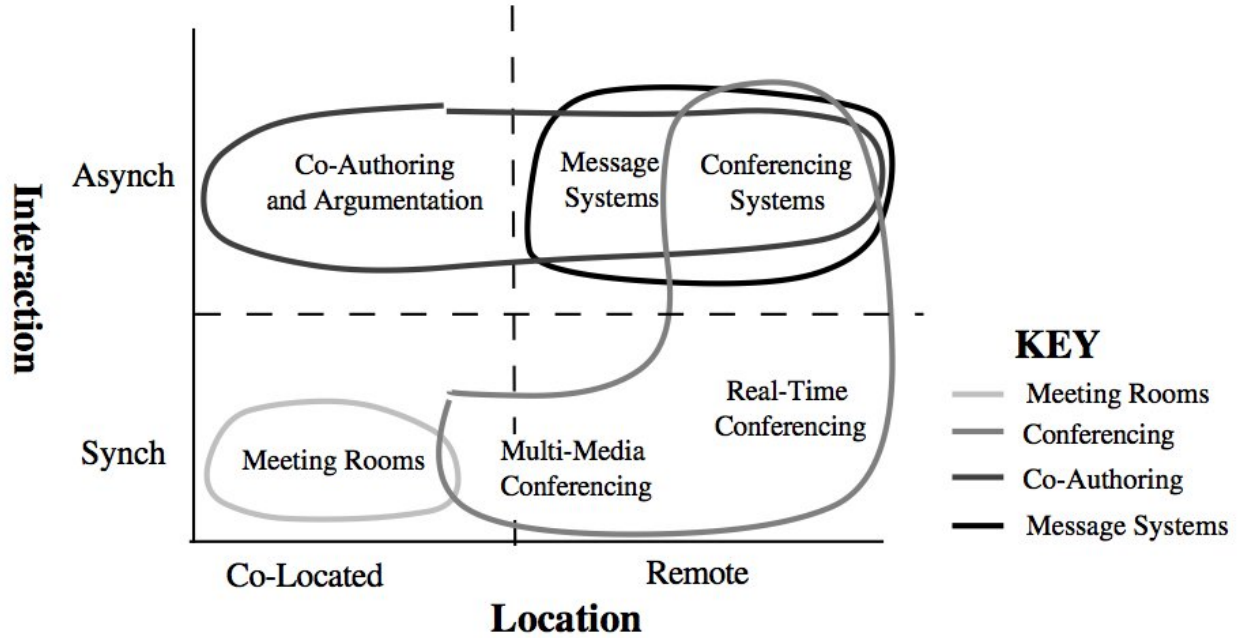


Figure 2.18: A classification space for CSCW systems (Rodden, 1991).

brief descriptions of them.

Table 2.3: Different scenarios of collaborative information processing.

User	Source	Time	Examples
Single	Single	–	Typical search
Single	Multiple	Asynchronous	Multi-source search
Single	Multiple	Synchronous	Meta-search
Multiple	Single	Asynchronous	Collaborative filtering, Collaborative navigation, Collaborative IR
Multiple	Single	Synchronous	Collaborative navigation, Collaborative IR
Multiple	Multiple	Asynchronous	Collaborative filtering, Collaborative IR
Multiple	Multiple	Synchronous	Collaborative navigation, Collaborative IR

- *Single user mode search*

This is a typical search scenario. A user issues a query to a search engine and

receives a ranked list. Relevance is found by considering various factors about individual documents, the whole collection, and links (Brin and Page, 1998; Kleinberg, 1999). Relevance feedback (Buckley, Salton, and Allen, 1994) and personalization (Teevan et al., 2005) are common ways to improve search in this mode.

- *Multi-source search*

No search engine has full coverage of the web (Sullivan, 2005). Issuing the same query to different search engines typically yields different sets of results.

- *Meta-searching*

Instead of a user issuing a query to different search engines, a system can do so simultaneously. It then combines the results obtained from a set of search engines, re-ranks them somehow and presents a single rank list to the user (Aslam and Montague, 2001). Examples include Dogpile and Clusty.

- *Collaborative filtering or recommender systems*

If there are multiple users using the same source for their information need, the source can keep track of what the users are looking for and what they are finding. Based on this tracking statistics, it can then make recommendations to other users who are also looking for the same or similar things. Amazon.com is an example of such a system, and there are plenty of other applications in use, some of which were mentioned in Section 2.3.

Several systems designed to provide co-browsing experience tend to be more about collaborative filtering than browsing due to their asynchronous nature and lack of two-way interaction among the participants. For instance, Wittenburg, Das, Hill, and Stead (1995) came up with the notion of *Group Asynchronous Browsing* (GAB) to provide tools for people to leverage the information hunting and

gathering activities of other people or group of people on the web. The authors created a server that collected and merged bookmark files of participating users and then served subsets of those merged bookmark files to either standard HTML client browsers or to a client built with the multiscale visualization tool Pad++. For the latter, they built a tool called WebWatch that could monitor URLs of interest and alert users when significant updates appear.

- *Collaborative IR*

This refers to multiple users working synchronously or asynchronously on single or multiple sources for the same information need. We will look at it in detail in Section 2.6.5.

- *Collaborative navigation*

This refers to multiple users browsing single or multiple sources together (co-located or remote, but synchronous). This will be explored further in Section 2.6.4.

2.6.3 Control, communication, and awareness in a CIS environment

Three components specific to group-work or collaboration that are highly predominant in the CIS or CSCW literature are control, communication, and awareness. In this section key definitions and related works for these components will be highlighted. Understanding their roles can also help us address various design issues with CIS systems.

Control

Rodden (1991) identified the value of control in CSCW systems and listed a number of projects with their corresponding schemes for implementing for control. For instance,

the COSMOS project (Wilbur and Young, 1988) had a formal structure to represent control in the system. They used roles to represent people or automata, and rules to represent the flow and processes. Roles of the people could be supervisor, processor, or analyst. Rules could be a condition that a process needs to satisfy in order to start or finish. Due to such a structure seen in projects like COSMOS, Rodden classified these control systems as procedural based systems.

Most of these systems were studied in office environments, where the subjects interacted with one another through personal conversations, group meetings, and phone calls. Several of the recommendations and findings of these studies were primarily based on observations.

To express control in a collaborative environment, early CSCW systems used various mechanisms to pass the messages around. These messages were often referred to as Structured Definition Language (SDL) messages. In the most basic sense, these were email messages that were sent back and forth among the participants of a collaborative project. However, for a collaborative project, an organization often needs more support than simply passing the information in messages. SDL provides this support by imposing a structure to these messages, and incorporating additional fields of information that can be used to filter and distribute messages appropriately.

For instance, Malone, Grant, Turbak, Brobst, and Cohen (1987) proposed the Information Lens framework, in which the messages carried additional information (some of which was automatically generated) that can later be used to filter and classify the messages to suit an individual's need in a group.

Let us see this by an example. Sam sends a message to circulate in the group, that he belongs to, asking for an opinion. At the time of sending this, Sam was not clear about who might be the right person in the group to inquire about that question, but since the system has additional information, such as user profiles and preferences, it

can use it to redirect and distribute the message to appropriate individuals. This way, Sam does not have to worry about looking for the right people, and the receivers do not have to worry about getting the messages that are not right for them, even though they are useful for the group as a whole.

Looking at the above scheme with a different perspective, we are distributing the control between humans and automatons involved in the whole group process. Instead of explicitly deciding by himself who should receive the message, Sam is letting the system take charge of this process, thus relinquishing the control to the system. The system is driven by the rules that guide its decision-making. It is important to note that such a system is different than a traditional collaborative filtering system, where the system filters information based on similarities among the users. Here, the messages are filtered based on sender's intention (Sam choose to distribute his message this way), and receivers' intention to receive such messages that are relevant to them. Malone referred to such kind of filtering as *cognitive filtering*.

Later Malone extended the above framework to Object Lens (Malone and Lai, 1988), in which the participants could create not only the messages to pass the information around, but any kind of objects. Each of these objects would have similar structure imposed on them that could guide further control and distribution processes. Object Lens also let people create links among those objects formed. Malone pointed out that this was similar to hypertexts on the World Wide Web.

Communication

This is one of the most critical components of any collaboration. In fact, Rodden (1991) identified message or communication systems as the class of systems in CSCW that is most mature and most widely used.

Since the focus here is on CIS systems that allow its participants to engage in an

intentional and interactive collaboration, there must be a way for the participants to communicate with each other. What is interesting to note is that often, collaboration could begin by letting a group of users communicate with each other. For instance, Donath and Robertson (1994) presented a system that allows a user to know that others were currently viewing the same webpage and communicate with those people to initiate a possible collaboration or at least a co-browsing experience. Providing communication capabilities even in an environment that was not originally designed for carrying out collaboration is an interesting way of encouraging collaboration.

Using four multidisciplinary design situations in the USA and Europe, Sonnenwald (1996) came up with 13 communication roles. The author showed how these roles can support collaboration, among other aspects of information seeking process, such as knowledge exploration and integration, and task and project completion, by filtering and providing information and negotiating differences across organizational, task, discipline and personal boundaries.

Awareness

Awareness is one of the most important issues that is identified and addressed in the CSCW literature. A clear definition and a methodology for providing awareness, though, are lacking. One of the often-asked questions about awareness in CSCW is “*awareness of what?*” Schmidt (2002, p. 288) argued that we should talk about awareness not as a separate entity, but as somebody’s being aware of some particular occurrence. In other words, the term ‘awareness’ is only meaningful if it refers to a person’s awareness *of* something. Heath, Svensson, Hindmarsh, Luff, and Lehn (2002) suggested that awareness is not simply a ‘state of mind’ or a ‘cognitive ability’, but rather a feature of practical action which is systematically accomplished within developing course of everyday activities.

Several related terms and definitions are used in the CSCW literature to refer to awareness in collaborative projects. For instance, Dourish and Bellotti (1992, p.107) defined awareness as *“an understanding of the activities of others, which provides a context for your own activity”*. Dourish and Bly (1992, p.541) suggested the following definition for awareness: *“Awareness involves knowing who is “around”, what activities are occurring, who is talking with whom; it provides a view of one another in the daily work environments. Awareness may lead to informal interactions, spontaneous connections, and the development of shared cultures - all important aspects of maintaining working relationships which are denied to groups distributed across multiple sites.”*

A set of theories and models for understanding and providing awareness emerged in the early works reported in the CSCW literature. Gaver (1991) argued that an intense sharing of awareness characterizes *focused collaboration* in which people work closely together on a shared goal. He further claimed that less awareness is needed for division of labor, and that more casual awareness can lead to serendipitous communication, which can turn into collaboration. He proposed a *general awareness* model that incorporates and supports all of such activities (Figure 2.19).

Bly, Harrison, and Irwin (1993, p. 29) also identified the importance of such general awareness by saying *“When groups are geographically distributed, it is particularly important not to neglect the need for informal interactions, spontaneous conversations, and even general awareness of people and events at other sites.”*

Some of the early works reported using ethnographic field studies in CSCW (e.g., Harper, Hughes, and Shapiro, 1989b; Harper, Hughes, and Shapiro, 1989a; Heath and Luff, 1991) identified the need to seamlessly align and integrate the activities of the participants of a collaborative project. While they did not refer to it as ‘awareness’, soon, the term ‘awareness’ was adopted to address such practices that support connecting collaborators without the activities of asking, suggesting, requesting, ordering, or

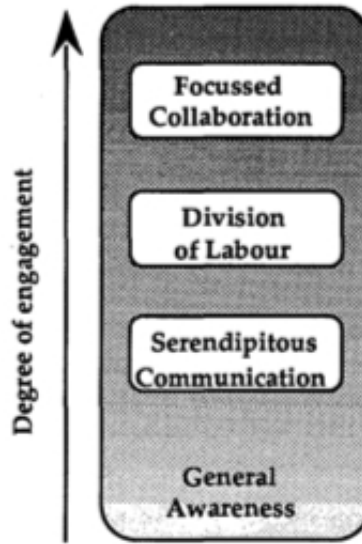


Figure 2.19: General awareness supporting serendipitous communication, division of labor, and focused collaboration (Gaver, 1991).

reminding.

Several works argued that providing audio-video communication channel could suffice for awareness (Mantei, Baecker, Sellen, Buxton, and Milligan, 1991; Gaver, Moran, MacLean, Lvstrand, Dourish, Carter, and Buxton, 1992). However, the use of communication as a substitute for awareness turned out to be very limited in its applicability (Gaver, 1992). Another line of research focused on providing awareness using *computational environments* based on ‘event propagation mechanisms’ for collecting, disseminating, and integrating information concerning collaborative activities. Some of the notable works in this stream of research include awareness models based on a spatial metaphor by Rodden (1996), and Sandor, Bogdan, and Bowers (1997). Related works in this direction were discussed in Section 2.6.1.

As we saw above, there are several ways of defining and implementing awareness. Various research projects have used their own taxonomy and interpretation of awareness for creating frameworks and systems. For instance, Gutwin and Greenberg (2002) classified awareness in two types: *situational*, and *workspace*, and suggested that situational

awareness underlies the idea of workspace awareness in groupware systems. Their definition of workspace awareness included how people interact with the workspace, rather than just awareness of the workspace itself. Simone and Bandini (2002) identified two kinds of awareness: *by-product awareness* that is generated in the course of the activities people must do in order to accomplish their collaborative tasks; and *add-on awareness* that is the outcome of an additional activity, which is a cost for the collaborators to what they must do and is discretionary in that it depends on collaborators' evaluation of the contingent situation. Chalmers (2002), likewise, divided the awareness in two kinds: awareness of *people*, and of *information artifacts*. He suggested implementing activity-centered awareness tool, in that it focuses on presenting the ongoing appearance and activity of people.

For the purpose of this dissertation, a more comprehensive and well-accepted taxonomy of awareness, which addresses four kinds of awareness (Liechti and Sumi, 2002) as listed below, will be used.

1. *Group awareness*. This kind of awareness includes providing information to each group member about the status and activities of the other collaborators at a given time.
2. *Workspace awareness*. This refers to a common workspace that the group members share and where they can bring and discuss their findings, and create a common product.
3. *Contextual awareness*. This type of awareness relates to the application domain, rather than the users. Here, we want to identify what content is useful for the group, and what the goals are for the current project.
4. *Peripheral awareness*. This relates to the kind of information that has resulted from personal and the group's collective history, and should be kept separate from

what a participant is currently viewing or doing.

Different CIS systems have different ways of providing awareness to the collaborators depending on the domain and the kind of application. Take, for example, Ariadne (Twidale, Nichols, and Paice, 1995), developed to support the collaborative learning of database browsing skills. To facilitate complex browsing processes in collaboration, Ariadne presents a visualization of the search process (Figure 2.20). This visualization consists of thumbnails of screens, looking like playing cards, which represented command-output pairs. Any such card can be expanded to reveal its details. The support for awareness, in this case, is driven by the specific domain (library) and application (catalogue search).

SearchTogether (Morris and Horvitz, 2007), on the other hand, was based on information seeking (application) on the web (domain). It instantiates awareness in several ways, one of which is per-user query histories. This is done by showing each group member's screen name, and his/her photo and queries in the "Query Awareness" region (Figure 2.21). The access to the query histories is immediate and interactive, as clicking on a query brings back the results of that query from when it was executed. The authors identified query awareness as a very important feature in collaborative searching, which allows group members to not only share their query terms, but also learn better query formulation techniques from one another. Another component of SearchTogether that facilitates awareness is the display of page-specific metadata. This region includes several pieces of information about the displayed page, including group members who viewed the given page, and their comments and ratings. The authors claim that such visitation information can help one either to choose to avoid a page already visited by someone in the group to reduce the duplication of efforts, or perhaps choose to visit such pages, as they provide a sign of promising leads as indicated by the presence of comments and/or ratings.

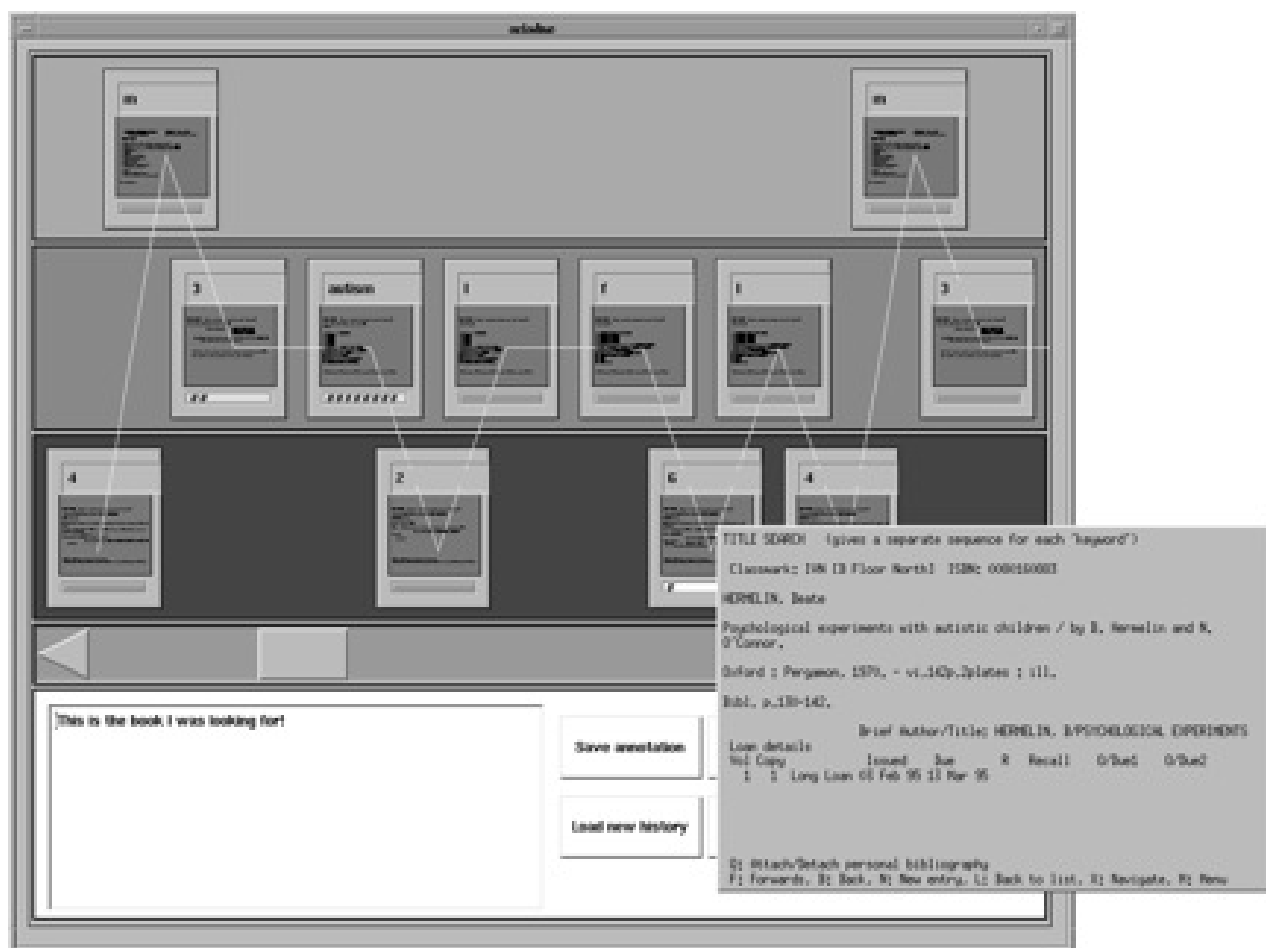


Figure 2.20: A search visualization in Ariadne (Twidale et al., 1995).



Figure 2.21: Screenshot of SearchTogether. (a) integrated messaging, (b) query awareness, (c) current results, (d) recommendation queue, (e)(f)(g) search buttons, (h) page-specific metadata, (i) toolbar, (j) browser. (Morris and Horvitz, 2007).

Importance of control, communication, and awareness in CIS systems

The findings from empirical observations and other studies of usability testing relating to control, communication, and awareness inform us that an effective CIS system should have the following attributes.

1. A flexible mechanism to incorporate structured message passing.
2. A way of facilitating control among the participants as well as with automaton components.
3. Facilities to present awareness of various objects, processes, and people at any given time to everyone in the group.

While these attributes are derived from general collaborative systems, they apply to CIS systems too. Let us explore how such attributes can fit in a CIS system.

Several CIS systems, such as SearchTogether (Morris and Horvitz, 2007) incorporate support for chat or IM. Such a support is crucial as the group members need a way to communicate with each other. Some works have also tried to provide other sorts of communication channel in a collaborative workspace, such as audio chat, video conferencing, and bulletin board support. While chat is an obvious choice for synchronous communication, email still prevails when it comes to providing asynchronous communication. In fact, Krutz claimed (in the 80s) that the most successful CSCW application was email. Recently, Morris (2008) found from a survey of knowledge workers that email is still one of the most used methods of communicating while working on a collaborative project. Given the importance of email, and the level of familiarity and comfort that most people have with it, an effective CIS system should provide support for passing such messages among the participants during collaboration.

In addition to this, we need to have some kind of structure imposed on the messages passed to incorporate additional information, such as time stamps, tags, and associated

processes. Such structure and information can be helpful in distributing the messages with some sort of filtering and/or following rules and roles of a system. For instance, Pickens, Golovchinsky, Shah, Qvarfordt, and Back (2008) demonstrated a collaborative video search system where one of the participants was responsible for issuing queries (prospector), and the other participant was responsible for going through the results looking for relevant information (miner). In its most basic version, this system had pre-defined roles and these roles followed a fixed set of rules. However, with the ability to have structured messages with appropriate information, we can have more flexible roles with dynamic distribution of control among the participants and the system.

Finally, providing awareness is highly important for a CIS system. Since the users of a CIS system will be working with different sources, documents, queries, snippets, and annotations of varying kind, we need to keep everyone in the group aware of all such objects as they are collected and modified. In addition to this, it is important to show various attributes associated with an object. For instance, it is useful to indicate on the interface that a document has already been viewed.

Several systems supporting collaboration have identified the above issues (control, communication, and awareness) as critical to their design. For instance, Farooq, Ganoie, Carroll, and Giles (2009) presented a collaborative design for CiteSeer, a search engine and digital library of research literature in the computer and information science disciplines. Based on a survey and follow-up interviews with CiteSeer users, the authors presented four novel implications for designing the CiteSeer collaboratory: (1) visualize query-based social networks to identify scholarly communities of interest, (2) provide online collaborative tool support for upstream stages of scientific collaboration, (3) support activity awareness for staying cognizant of online scientific activities, and (4) use notification systems to convey scientific activity awareness.

Let us turn our attention to some of the systems developed to facilitate CIS. These

systems can be classified in two broad categories: co-browsing or social navigation, and collaborative IR.

2.6.4 Co-browsing or collaborative/social navigation

Co-browsing or social navigation is a process of allowing a set of participants navigate or browse, and share information with a possible intermediate interface. Root (1988) introduced the idea of social browsing to support distributed cooperative work with unplanned and informal social interaction. He described a “social interface”, which provided direct, low-cost access to other people through the use of multimedia communications channels. The design of his conceptual system, called CRUISER,¹⁶ incorporated three basic concepts: social browsing, a virtual workspace, and interaction protocols. His premise was that by integrating all of our digital media into a richly interconnected workspace, we could significantly extend and enrich the available context of our workgroup activities.

Root’s idea of facilitating informal and effortless interaction among a group of people was carried over later by Donath and Robertson (1994) with *The Social Web* that allowed a user to know that others were currently viewing the same webpage and communicate with those people. They believed that users accessing the same page are likely to be in search of the same type of information and share similar interests. Providing them with the ability to communicate with each other can facilitate information searches and help foster community.

Cabri, Leonardi, and Zambonelli (1999) presented a system for synchronous cooperative browsing that permitted users within a workgroup to share information and

¹⁶ “Cruising” was the stereotypical teenage activity of the 50’s and early 60’s. The term refers to the practice of piling into somebody’s car and visiting the chain of gathering places frequented by other peer group members, or simply driving around in search of almost any sort of social encounter (Lucas, 1973).

cooperate toward a common goal. This was done using a proxy without changing the browsers on user ends. Gerosa et al. (2004) presented a similar idea of proxy-based co-browsing with the application of e-learning. They called this *Symmetric Synchronous Collaborative Navigation*, a form of social navigation, where users virtually share a web browser. They presented a symmetric, proxy-based architecture implemented without the need for a special browser or other software. Once again, the motivation behind such light-weight interfaces was to allow the users to emerge into a collaborative environment with as little effort as possible. Esenther (2002) emphasized having a light-weight real-time collaborative web browsing service and providing an instant co-browsing facility. Their system was targeted to casual (non-technical) users and allowed remote participants to easily synchronize pointing, scrolling and browsing of uploaded content in their web browsers.

Another example of collaborative browsing application is *AntWorld* (Menkov, Neu, and Shi, 2000), a tool developed to make it easier for the members of a common-interest user group to collaborate in searching the web. *AntWorld* harnesses the expertise of the members of a common interest group as displayed by their evaluation of documents encountered while searching. It stores users judgments about the documents they find and uses this information to guide other users to pages they may find useful.

Sometimes it is not just the webpages that people want to browse and share, but other objects such as bookmarks. Keller, Wolfe, Chen, Rabinowitz, and Mathe (1997) presented *WebTagger*, a social bookmarking service similar to del.icio.us, which allowed a group of users to tag and share webpages. *WebTagger* enables users to supply feedback on the utility of the resources that they bookmarked relative to their information needs, and provides dynamically-updated ranking of resources based on incremental user feedback.

Several other systems used their own interfaces rather than relying on a web browser.

For instance, *GroupWeb* (Greenberg and Roseman, 1996) is a browser that allows group members to visually share and navigate World Wide webpages in real time. Its groupware features include document and view slaving for synchronizing information sharing, telepointers for enacting gestures, and “what you see is what I see” views to handle display differences. *GroupWeb* also incorporated a groupware text editor that lets groups create and attach annotations to pages. Similarly, *GroupScape* (Graham, 1997) was a multiuser HTML browser to support synchronous groupware applications and browsing of HTML documents on the web.

Yet another architecture to support multiuser browsing is *CoVitesse* (Laurillau and Nigay, 2002), a groupware interface that enables collaborative navigation on the web based on a collaborative task model. This system represented users navigating collaboratively in an information space made of results of a query submitted to a search engine. In contrast to these systems, which are primarily designed for remotely located participants, *CoSearch* (Amershi and Morris, 2008) is implemented to provide multi-device support for collaborative browsing among co-located participants.

Some of the applications allow the users of that system to play different roles during their social or collaborative browsing for information. For instance, Pickens et al. (2008) proposed the roles of *Prospector* and *Miner* in a collaborative video search environment, the former one responsible for seeking out various areas where relevant information could be found, and the latter one responsible for digging deeper in a given sub-domain with high likelihood or useful information. A collaborative navigation system proposed by Gerosa et al. (2004) had the provision where each user could take the lead and guide others in visiting websites. However, Aneiros and Estivill-Castro (2003) advocated against controlled co-browsing where one user guides the browsing process for the others (what they referred as the *master/slave model*) and proposed to use a model with unconstrained collaborative web browsing. They argued that such

unconstrained collaborative web navigation is essential to allow natural information flow among multiple users.

2.6.5 Collaborative IR

The discussion will now be narrowed down to those scenarios in collaborative setup where the goal is to seek information together for a common information need. As discussed earlier, if/when the problem of IR is difficult to solve, a carefully executed collaboration can help. Smyth, Balfe, Briggs, Coyle, and Freyne (2003) argued that one way of making it possible to connect users to the information that is difficult to find is to incorporate collaboration in the search phase of an information seeking process. They showed how collaborative search can act as a front-end for existing search engines and re-rank results based on the learned preferences of a community of users. They attempted to demonstrate this concept by implementing the I-Spy system (Freyne, Smyth, Coyle, Balfe, and Brigg, 2004). I-Spy captures the queries and the related results for a given workgroup and uses that information to provide filtered, and presumably more relevant, information to the user. Thus, I-Spy acts more as a collaborative filtering process than as synchronous collaborative searching.

While I-Spy attempts to extend content-based filtering techniques by incorporating communities, several collaborative IR systems are developed by extending a traditional IR model to incorporate multiple users. However, such extension is often ineffective or non-trivial. For instance, Hyldegard (2006), with her studies of information seeking and retrieval in a group-based education setting, found that even though people in a collaborative group to some extent demonstrated similar cognitive experiences as the individuals in Kuhlthau's Information Search Process (ISP) model (Kuhlthau, 2005), these experiences did not only result from information seeking activities, but

also from work-task activities and intragroup interactions. Her further work also indicated (Hyldegard, 2009) that group based problem solving is a dynamic process that shifts between a group perspective and an individual perspective. Such a finding calls for a thorough investigation into collaborative information seeking that is not simply an extension of a traditional IR system for multiple users. As Olson, Olson, Carter, and Storrosten (1992, p.347) suggested, *“The development of schemes to support group work, whether behavioral methods or new technologies like groupware, should be based on detailed knowledge about how groups work, what they do well, and what they have trouble with.”*

Unlike co-browsing, where the applications are aimed toward web browsing, works on collaborative IR are often focused on specialized domains for searching. For instance, Twidale and Nichols (1996) presented the Ariadne system, which allowed a user to collaborate with an information expert remotely and synchronously over a library catalogue. The idea behind Ariadne was to allow the patron (naive user) to collaborate with a reference librarian (search expert) for an information need in a library situation. The authors identified the importance of supporting social aspects of searching for information and showed how it can be addressed using their system. However, Ariadne did not have support for asynchronous collaboration.

Morris and Horvitz (2007) presented the SearchTogether system that allowed a group of remote users to collaborate synchronously or asynchronously. This system was based on supporting awareness, division of labor, and persistence for collaboration. Their rationale for facilitating awareness was that it could enable light-weight collaboration by reducing overhead involved in explicitly asking other group members to provide that information. Awareness was provided using per-user query histories, page-specific metadata, and annotations. Division of labor was implemented using

integrated IM as well as a recommendation mechanism, by which a participant can recommend a page to another participant. SearchTogether also provided “Split Search” and “Multi-Engine Search” options for automatic division of labor. Finally, persistence was implemented by storing not only all session state, but also automatically creating a shared artifact that summarizes the findings of a collaborative search.

MUSE (Krishnappa, 2005) supports synchronous, remote collaboration between two people searching a medical database. MUSE lets its users perform standard single-user searches, with a provision of chat and the ability to share metadata about the current database results with the other user. S^3 (Morris and Horvitz, 2007) is not so much of a CIS system, but it has an important component of being able to share the retrieved results asynchronously among a set of users.

A stream of research came out of the CIR group at University of Washington (Washington, 2007), studying the situations where members of a work-team are seeking, searching, and using information collaboratively and showing how such a process can be realized in a multi-team setting. This started with (Fidel, Bruce, Dumais, Grudin, Poltrock, and Pejtersen, 1999), where the authors defined collaborative IR (CIR) “*as 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.*” They employed a cognitive work analysis framework to guide a field study examining social, organizational, cognitive, and individual characteristics of information seekers, and then focusing to address collaborative situations (Fidel, Bruce, Pejtersen, Dumais, Grudin, and Poltrock, 2000b). From their studies involving two design teams working in collaboration, they found (Bruce, Fidel, Pejtersen, Dmais, Grudin, and Poltrock, 2003) that (1) the nature of the task and the structure and the culture of the organization in which tasks are performed are important factors that determine CIR behavior, and (2) not all information behavior takes place collaboratively even in teams that carry

out CIR. In their further work on this line, the authors found (Pollock, Grudin, Dumais, Fidel, Bruce, and Pejtersen, 2003) that (1) any information retrieval activity (identifying information needs, formulating queries, retrieving information, evaluating it, and applying it to address the need) may be performed by an individual on behalf of the team, by an ad-hoc group, or by the team working together in a meeting, and (2) technologies intended to support teamwork could be more effective by recognizing and supporting collaboration in the activities that comprise information retrieval and their coordination. This suggests that a successful CIR/CIS system should not try to lock the users down in a certain kind of framework imposed by that system; it should rather let the participants choose their own way of collaborating, and provide enough support for carrying out that collaboration.

The efforts of connecting multiple users for information seeking (retrieval or browsing) continue to produce systems either by reinventing the wheel of traditional IR, or by extending existing IR systems to accommodate more than one user. None of these systems have been adopted widely in practice. Several reasons can be found for the lack of wider visibility of collaborative systems, among which are the cognitive load involved in using these systems, the learning curve to start using these environments, and a lack of proper integration of information seeking to other parts of the collaboration.

2.6.6 Realization of a collaborative environment

There are several ways in which a collaborative environment can be realized. We are familiar with using tools such as telephones and email for doing collaboration with remotely located users asynchronously or synchronously. However, these tools are not specifically designed to do collaboration. If we take a step back and analyze what we really need to do an effective collaboration, we may come up with a different set of

tools. To understand the issues in implementing a collaborative system, three important aspects can be identified: processes, content, and devices. Following is a brief description of each of these aspects along with related works.

Processes: In its most basic case, a collaborative environment lets the collaborators divide up the workload and/or combine their inputs somehow. Several of the realizations of a collaborative environment have primarily focused on the system side. These works typically present some algorithmic way of combining multiple instances of search requests, result lists, or other interactions from different users to perform implicit “collaboration”. For instance, a good deal of work in implementing a collaborative search system has been done around reformulating search requests of a user based on other users’ search requests on the same/similar search goals. Fu, Kelly, and Shah (2007) showed how different queries from a set of users for the same information goal can be combined, calling them “collaborative queries”, for better retrieval performance.

Content: A simple way of taking advantage of collaboration is dividing the content among the users for viewing, judging, or manipulating. With *WebSplitter*, Han, Perret, and Naghshineh (2000) demonstrated how a unified XML framework can support multi-device and multi-user web browsing. *WebSplitter* splits a requested webpage and delivers the appropriate partial view of each page to each user, or more accurately to each user’s set of devices. Multiple users can participate in the same browsing session, as in traditional conferencing groupware. Similarly, Maekawa, Hara, and Nishio (2006) developed a page partitioning method for collaborative browsing, which divides a webpage into multiple components. They also designed and implemented a collaborative web browsing system in which users can search and browse their target information by discussing and watching partial pages displayed on multiple devices. As a part of

their SearchTogether system, Morris and Horvitz (2007) incorporated a feature that lets users dynamically distribute the webpages among themselves. The authors identified such a division of labor as one of the appeals of their collaborative IR system.

Devices: Typically, in a CIS environment, by system, we are referring to computers, but several works have tried to extend information access and distribution to other forms of devices to enable collaboration among the users in a variety of work places. For instance, Maekawa, Hara, and Nishio (2006) presented a collaborative web browsing system in a mobile computing environment. Their motivation for using collaboration in the mobile device environment was to overcome the issue of low functionality that restricts the services provided for mobile users. Rutgers's DISCIPLE (Distributed System for Collaborative Information Processing and LEarning) project (Rutgers, 2009) is another example of introducing collaborative processes to mobile users. The key objective of the DISCIPLE project has been to develop an advanced groupware design that enables interactive collaboration in the context of the task at hand. Amershi and Morris (2008) presented *CoSearch* - a collaborative browsing interface to be used on computers, and introduced *CoSearchMobile*, designed to provide similar functionalities on mobile devices. *CoSearch* system leverages readily available devices such as mobile phones and extra mice to facilitate collaborative browsing among co-located users.

Blackwell, Stringer, Toye, and Rode (2004) described a tangible interface for collaborative IR. The purpose of this interface was to allow multiple users to interact simultaneously to refine a query. Morris, Paepcke, and Winograd (2006) presented *TeamSearch* system, which used an interactive table for a small group of co-located participants in searching for digital images to use in a report. Mitsubishi Electric Research Lab (MERL) has developed DiamondTouch (Smeaton, Lee, Foley, and Givney, 2006b), an interface device that supports direct user collaboration on a tabletop. Such

an interactive tabletop is ideal for multimedia searches in collaboration. Smeaton, Foley, Gurrin, Lee, and Givney (2006a) and Smeaton, Lee, Foley, Givney, and Gurrin (2006c) reported video searching in collaboration using such DiamondTouch interactive tabletop devices. Among other things, the authors found about a 10% increase in the level of user-interaction as the users moved from their first search to the last one with DiamondTouch.

2.7 Evaluation

Evaluating a CIS environment can be a huge challenge due to its complex design that involves a set of users, integrated systems, and a variety of interactions. One can evaluate a CIS system using typical measures of IR. However, as discussed before, information seeking is not merely about retrieving information, and thus, evaluating a CIS system with its retrieval effectiveness may not be sufficient. While traditional IR evaluations can still be used to measure the retrieval performance of a collaborative filtering system, just as Smyth, Balfe, Boydell, Bradley, Briggs, Coyle, and Freyne (2005) did, we need additional measures for CIS systems.

Baeza-Yates and Pino (1997) presented some initial work on trying to come up with a measure that can extend the evaluation of a single-user IR system for a collaborative environment. While this was based on the retrieval performance, Aneiros and Estivill-Castro (2005) came up with the proposal of evaluating the “goodness” of a collaborative system with usability. In addition, Baeza-Yates and Pino (1997) treated the performance of a group as the summation of the performances of the individuals in the group. While this may work for simple information seeking and retrieval, we can imagine situations in which this is not true. For instance, if two people working together can find twice as much information as either of them working independently, was that a *good* thing? How about the amount of time they spent cumulatively? The

participants may not be able to find twice as many results, but what if they achieved *better understanding* of the problem or the information due to working in collaboration? Then there are other factors, such as *engagement*, *social interactions*, and *social capital*, which may be important depending upon the application, but are usually not looked at in non-interactive or a single-user IR evaluations.

The majority of the work reported in the literature that has attempted to evaluate the effectiveness of a collaborative system has looked at the usability of the collaborative interface. For instance, Morris and Horvitz (2007) tested their SearchTogether system with a user study to evaluate how users utilize various tools offered in their interface and how those tools affect the act of collaboration. The authors used seven pairs of users and let each pair choose their topic of mutual interest to work with. The evaluation was based on the log, observations, and questionnaire data. While they showed the effectiveness of their interface in letting people search together, there was no evaluation of learning that took place in the group due to collaboration. Laurillau and Nigay (2002) demonstrated how multiple users can navigate the web in a collaborative environment with their *CoVitesse* system. They presented evaluations for the user interface as well as various network-related parameters. However, no clear understanding of the effects on the retrieval performance was reported. Aneiros and Estivill-Castro (2005) presented a questionnaire to the participants of their user study to evaluate the usability of their Group Unified History (GUH) tool. Typical questions on their questionnaire were “*how difficult was it to interpret the user identity symbols used in the tool?*” and “*did you visit any websites found by your team/peers using the group history?*”

Smyth et al. (2003) tested their I-Spy system with leave-one-out evaluation methodology. From 20 users, they left one user as a *testing user* and used the other 19 users as the *training users*. The relevancy results of the training users were used to populate I-Spy’s hit matrix (detail given earlier) and the results of each query were re-ranked

using I-Spy's relevancy metric. Then they counted the number of those results listed as relevant by the test user for various result-list sizes and finally, they made the equivalent relevancy measurements by analyzing the results produced by the untrained version of I-Spy to serve as a baseline.

Some of the application designers also let "real" users use their systems and evaluated the effectiveness of their system from these users' feedback and/or their success in solving their "real" problems with it. For instance, Twidale, Nichols and Paice (1995) invited volunteers to bring a problem that they already have to solve. Students from a wide range of academic backgrounds (including Psychology, Computing, Women's Studies, Chemistry, Religious studies and Environmental Science) used their Ariadne system. The typical case was that they were about to write an extended essay, dissertation or group project and needed to do a literature search. The testing informed the iterative development of the system.

Prekop (2002) presented a qualitative way of evaluating collaborative information seeking studies. He proposed this by measuring *information seeking patterns*. These patterns describe prototypical actions, interactions, and behaviors performed by participants in a collaborative endeavor. The three patterns that the author described were *information seeking by recommendation*, *direct questioning*, and *advertising information paths*. On the similar line of studying the participants by analyzing their behavior and patterns, Olson et al. (1992) studied 10 design meetings from four projects in two organizations. The meetings were videotaped, transcribed, and then analyzed using a coding scheme that looked at participants' problem solving and the activities they used to coordinate and manage themselves. The authors also analyzed the structure of their design arguments. The authors claimed that the coding schemes developed may be useful for a wide range of problem-solving meetings other than design.

Wilson and schraefel (2008) analyzed an evaluation framework for information seeking interfaces in terms of its applicability to collaborative search software. Extending Bates' tactics model (Bates, 1979) and Belkin's model of users (Belkin et al., 1993), they showed that the framework can be just as easily applied to collaborative search interactions as individual information seeking software, but pointed out that there are additional considerations about the individual's involvement within a group that must be maintained as the assessment is carried out.

These efforts of evaluating various factors in CIS can be summarized as measuring (1) retrieval performance of the system, (2) effectiveness of the interface in facilitating collaboration, and (3) user satisfaction and involvement. Despite these efforts, there is still a lack of clarity and methods in evaluating CIS environments that can measure factors such as learning, user engagement, and group performance. Given this, the research reported here can provide a valuable contribution with proposals and demonstrations of various evaluation metrics for collaborative systems.

2.8 Conclusion

A review of literature from five major domains that relate to Collaborative Information Seeking (CIS) was presented in this chapter. This section summarizes the important lessons learned from this review.

While exploring a general notion of collaboration and the motivations behind it, it was discovered that there is a variety of definitions of "collaboration" in the literature, often using it interchangeably with "coordination" and "cooperation". From the review of various works, an attempt was made to establish a working definition of collaboration, which involves a group of participants intentionally working together in an interactive manner for a common goal. It was clarified that the focus here is on facilitating such collaboration in information seeking domains that are mediated by a

system. A number of models for information seeking were then discussed and a model that is based on information access and organization was proposed. A review of a highly related field, information filtering, then followed. It was shown that some of the approaches in this field could be useful to CIS, and more importantly, how the proposed notion of collaboration is different than the one employed in those collaborative filtering systems. Since the interest here is in developing highly interactive systems, it was essential to look at various models of user and system interactions. With the review of such models, an attempt was made to classify various systems and situations according to the degree of involvement and explicitness by both the systems and the humans. This made it possible to orient the proposed research with respect to existing interactive environments. Finally, considering the importance of the social aspect for CIS, a number of works from the field of social networking were reviewed. It was pointed out that weaker ties observed in many social networks can be extended to create stronger connections leading to collaborations.

In the final phase of this review, a number of systems that address various issues of CIS and implement such environments for searching or browsing information in collaboration were discussed. It was also pointed out that the efforts to evaluate CIS systems or some of their aspects are challenging and under developed. To summarize, following key points were derived from the review of the literature.

1. Collaboration involves people working together for a common goal or solution; simply working together or interacting is not enough.
2. Collaboration is intentional and interactive.
3. The value of collaboration in information seeking depends on the kind of task; people may not find it useful to collaborate in simple known-item, fact-finding tasks.

4. Collaboration may help an individual participant to achieve what he could not while working in solitude.
5. A careful collaboration can help the team produce something that is more than the sum of individual participants' contributions.
6. Doing collaboration involves certain overhead and additional cognitive load. We need to consider this factor when evaluating a CIS environment.
7. Information exchange and filtering may be necessary conditions for collaboration, but they are not sufficient conditions.
8. Understanding derived from social networking about how people communicate and work with each other is very useful in realizing CIS environments.
9. While collaborating over an information seeking process, collaboration among the users can occur at various levels: (1) while formulating an information request, (2) while obtaining results, and (3) while organizing and using the results. A CIS system should support all of these levels for an effective collaboration.
10. Appropriate support for control, communication, and awareness is essential for a successful collaborative system. Provision of proper awareness can help the participants be more efficient and engaged in the collaboration.
11. The majority of CIS systems are implemented assuming it is enough to extend a single-user IR system to accommodate multiple users. A deeper understanding of how collaboration as well as information seeking works is required.
12. Often traditional IR or HCI measures are used to evaluate various parameters in a CIS environment. While they may serve for studying individual factors, holistic understanding of an environment is often not possible by the combination of such measures.

Having reviewed relevant literature, it is clear that there are still several pieces that are missing in the field of CIS. Some of them are as following.

- We have a fairly good understanding of why people collaborate. However, these motivations are often not identified in the context of situations in which collaboration occurs. Often, even if collaboration can be useful, people do not see the value in collaborating. We do not know how to identify such situations and promote collaboration to people.
- The literature points us to a list of tools and methods that people use for collaborating. However, the relative merits of these tools and methods are not very clear. Often people collaborate using tools that were not specifically designed for collaboration. One could argue that we need specialized tools to support CIS, but we do not know what such tools may look like and how we could promote them to people without causing additional burden to them.
- The CSCW literature identifies three major issues in a CIS environment: control, communication, and awareness. Control is domain specific; communication is system specific; but awareness may depend on several factors, including task, distribution of responsibilities among the collaborators, roles of the collaborators, nature of the final product, need for privacy and sharing among the collaborators, and the nature of their collaboration (synchronous vs. asynchronous, co-located vs. remote). The issue of awareness is highly understudied in the literature and a good understanding of implementing support for awareness in a CIS system would add considerable value to CIS theory and practice.
- The literature provides several suggestions for evaluating a CIS system as well as users' performance while working with such a system. However, it still remains unclear what factors we should measure and how. This is likely to depend on

the domain of the application. For instance, for a time-bound recall-oriented task such as one reported in (Pickens et al., 2008), we can use relevance and efficiency as measures. But such metrics may not be appropriate for an education setting, where learning is probably a more important factor to measure. A taxonomy of evaluation metrics for different CIS situations is needed.

- The literature studied here includes a number of works done to understand people working with collaborative systems such as SearchTogether, and people’s behavior in online communities and social networking sites. A link that connects these two is missing. In other words, we do not know how we can leverage people’s engagement in social networking sites to promote collaborations, or support various social activities with collaborative systems.
- A more fundamental need exists for extending or augmenting single-person information seeking models to people seeking information in collaboration.

These issues are at the core of the CIS domain, and studying them could help us get closer to having a better understanding of people’s behavior in CIS environments, as well as designing better CIS systems. From these, the present dissertation focuses on the issue of awareness in CIS environments. To be precise, the dissertation investigates the ways to provide an appropriate support for awareness specific to certain kinds of CIS situations, and to evaluate and enhance various user and system related factors.

Despite the narrowly defined focus of this dissertation, it will inform many threads of further research. For instance, this research can inform the effectiveness of various tools in the context of tasks or situations. It will also help us understand various behavioral patterns of the user in a collaborative environment. Finally, this dissertation is important and timely as even though the value of collaboration in dealing with complex tasks is intuitively convincing, (1) the majority of IR systems today do not

facilitate collaboration directly, and (2) the concept of collaboration itself is not well-understood.

With this understanding, a series of investigations was conducted looking at instances, motivations, and methods that people use for collaboration, emphasizing information seeking. These investigations supported design specifications for a CIS system. The lessons learned from designing and developing such a system serves as the main contribution of this dissertation.

Chapter 3

Design Investigations and System Development

As discussed during the literature review, CIS is a complex and multi-disciplinary problem. Having identified some of the major issues and many works that have tried to address them, the focus of this dissertation will now be narrowed down to certain specific problems. One such issue is understanding instances, motivations, and methods for collaboration. In addition to this, a design for a CIS system, called Coagmento,¹ is sketched that was subsequently tested.

As Grudin (1994, p.93) pointed out, *“many expensive failures in developing and marketing software that is designed to support groups are not due to technical problems; they result from not understanding the unique demands this class of software imposes on developers and users.”* An exploration in designing and developing a CIS system, therefore, will be attempted by understanding existing practices for collaboration. As we saw in the literature review of collaboration (Section 2.1) and of CIS (Section 2.6), several studies have tried to explore motivations, costs, benefits, and methods for doing collaboration in general, and collaboratively seeking information in particular. In this

¹In Latin, Coagmento means working or joining together.

chapter, such understanding developed from those studies serves as a base for a qualitative analysis of user needs in collaboration. Based on these explorations, a preliminary CIS system is developed (Section 3.1) and tested using a series of design studies (Sections 3.3 to 3.5). This has resulted in a more functional and suitable CIS system for online information seeking in collaboration (Section 3.6). A timeline, showing how the studies and different versions of Coagmento took place, is shown in Figure 3.1.

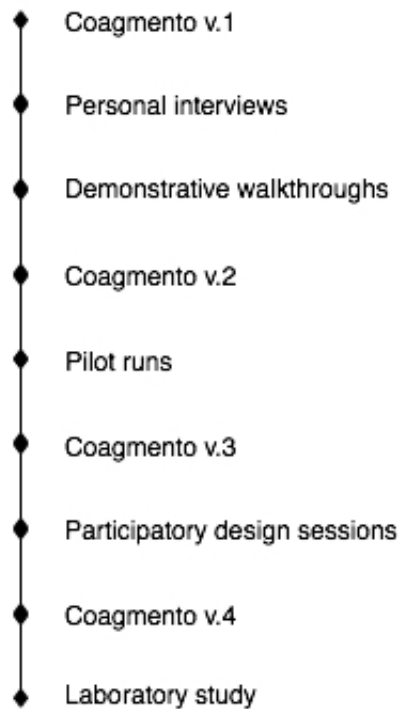


Figure 3.1: Timeline for Coagmento design and development, along with the studies.

3.1 Preliminary design of Coagmento

In order to come up with initial design specifications for Coagmento, two works are particularly helpful: one based on a general notion of collaboration, and the other more specific to designing a collaborative system. Surowiecki (2004) lists four conditions for

a successful collaboration: (1) diversity of opinion, (2) independence, (3) decentralization, and (4) aggregation. Morris and Horvitz (2007) presented the SearchTogether system based on supporting (1) awareness, (2) division of labor, and (3) persistence for collaboration. Based on these works, the following set of guidelines for designing a user-centered CIS system are inferred.

1. The system should provide an effective way for users to communicate with each other.
2. The system should allow (and encourage) each user to make individual contributions to the collaboration.
3. The system should coordinate user actions, information requests, and responses to support an active and interactive collaboration. This collaboration could be synchronous or asynchronous, and co-located or remote.
4. Users need to agree to and follow a set of rules to carry out a productive collaboration. For instance, if they have a disagreement on the relevancy of an information object, they should discuss and negotiate; they should arrive at a mutually agreeable solution rather than continuing to dispute it. The system needs to support discussion and negotiation processes among the users.
5. The system should provide a mechanism to let the users not only explore their individual differences, but also negotiate roles and responsibilities. There may be a situation in which one user leads the group and others follow (cooperate), but the real strength of collaboration lies in having the authority vested in the collective.

Using the above guidelines, a prototype system called Coagmento was developed that allows two people to work together for seeking information. Collaborators can work

synchronously or asynchronously, and they may be co-located or remotely connected. This version of Coagmento was designed to work with a specific collection rather than the open web. The main screen of Coagmento is shown in Figure 3.2. Let us now see how the above listed characteristics are realized in Coagmento.

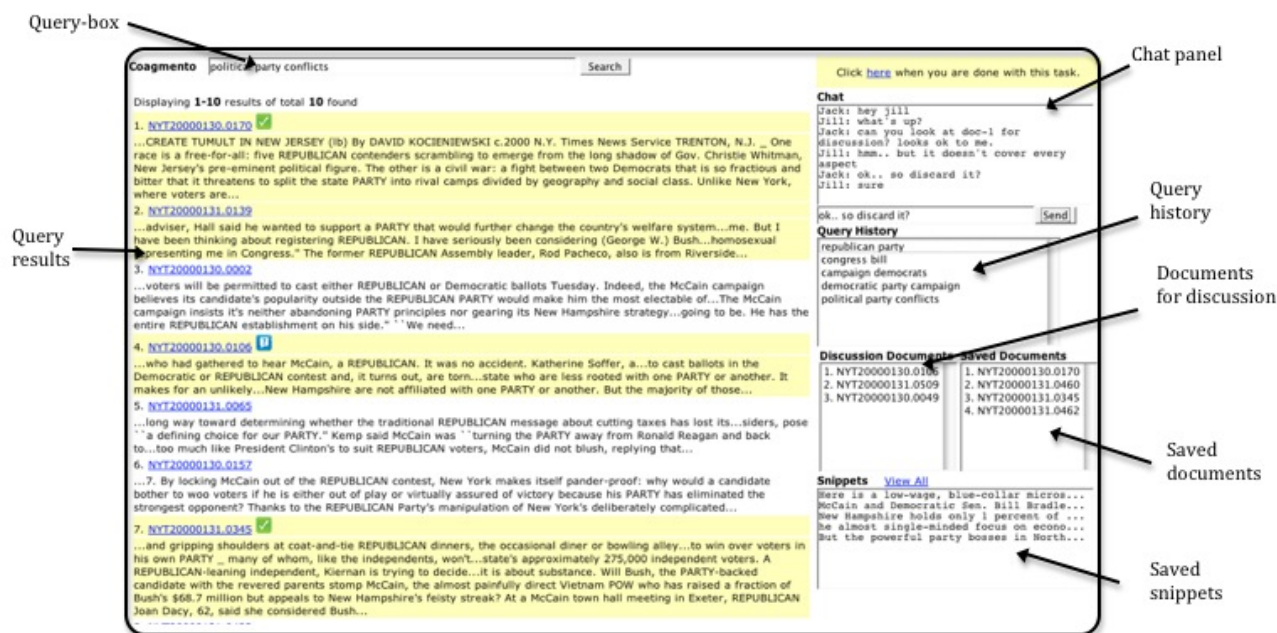


Figure 3.2: The main interface and its components in Coagmento.

1. As we can see, Coagmento includes a search interface, chat, and document space (the same space where the results are displayed in the figure), as well as various marking facilities (discussed later) - all in one place.
2. All of these tools are presented on a single interface and readily accessible. There is no need to switch to different windows or tabs as in SearchTogether.
3. Coagmento displays the partnership information and provides visual feedback based on one's partner's as well as one's own actions. For instance, if a document is already viewed by either of the persons in a pair, it will be highlighted anywhere it appears in a rank-list for both of them.

4. Coagmento keeps a log of all the queries used during a search session.² The list of these queries is presented on the interface. Unlike SearchTogether, clicking on a query executes fresh results, and not its history.
5. The users of Coagmento can save any document that they find useful or flag it to be discussed with their partners (Figure 3.3). Once again, these two lists are readily available on the interface and clicking on the name of a document there brings up that document to view.
6. If a user is working alone, he may not see much use in writing notes about everything that is saved as he may have an intuitive understanding about the relevance of those results. While working with a partner, on the other hand, the user may need to convey *what* and *why* aspects about the usefulness of a document. Coagmento allows a user to add notes to any document (Figure 3.3). Morris and Horvitz (2007) found such a feature useful, but they also realized that they needed a way for users to simply highlight and save portions of pages. Coagmento provides a way to ‘snip’ a passage while viewing a document that the user found useful (Figure 3.3). This is done by highlighting any text on any displayed page and clicking on the ‘Snip’ button on the toolbar.
7. Coagmento saves the state information. This means a user can leave a session and when he comes back, he will find the session as it was, with some possible updates in case his partner kept working while he was gone. This allows the users to collaborate in either synchronous or asynchronous mode. There is an indication on the interface to let a user know if his partner is online or not.

Based on the description above, it should be clear that in principle, Coagmento builds on the framework of other tools such as SearchTogether, and extends them

²The session can span several sittings or usage of the system throughout the project.

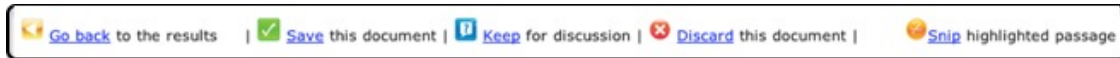


Figure 3.3: Toolbar provided on top of the document being viewed with Coagmento.

in certain ways. One aspect of SearchTogether that Coagmento does not implement fully is the division of labor. There are three ways in which this feature is realized in SearchTogether: (1) chat, (2) recommendations, and (3) split search. Coagmento has a chat feature, which can be used to talk about the distribution of the work. As far as the recommendations feature is concerned, the authors of SearchTogether found it underutilized. They concluded that rather than providing a “recommend” option, providing a “share this” option would allow a better way of sending pages back and forth. Coagmento does this through its “discuss this document” feature. For the SearchTogether system, it was found that the automatic division of labor features such as split search were not heavily used. The usefulness of such features needs further investigation.

3.2 Study-1: Personal interviews

Previously it was noted that there is a need to address even some of the very fundamental issues of collaboration, such as how and why people work together, what are the costs and benefits, and what methods and tools are used while collaborating. To investigate these issues and come up with design guidelines to build support for CIS, several people were interviewed individually regarding their past experiences with collaboration.

This study (Study-1) aimed to understand how people typically collaborate on a group project. In particular, the focus was on studying their motivations, the tools they use, and the methods they employ for such collaboration. For the study, several graduate students and faculty members, working in the field of social sciences, at UNC

Chapel Hill were contacted using a convenience sampling method. I knew these participants personally and knew that they had been involved in some collaborative work in the recent past. The interviews lasted about 30 to 45 minutes. The participants were not given any compensation.³

A total of 11 participants between age 25 and 58 were interviewed. All the participants were fairly to very experienced searchers, doing web searches every day. Almost all of them were involved in some collaborative projects on a day-to-day basis. While interviewing these participants, an attempt was made to elicit information about not only their past collaborations, but also other collaborative projects that they may have seen their colleagues, students, friends, or family members doing. Also, instead of asking about respondents' collaboration for information seeking tasks only, the questions were formulated looking at a larger context, focusing on the situations that got them into collaboration, how they worked together, and what they thought about the process in retrospect. The interview guide for these interviews is given in Appendix B.

Morris (2008) explored similar questions using a survey. She surveyed 204 knowledge workers at a large technology company to investigate collaborative web search practices that people employ. Her survey included questions to determine whether people need and/or want to collaborate while searching the web, and if they do, what strategies they employ to collaborate given that such an activity is not explicitly supported by current search interfaces.

The study reported here was much more preliminary and different in two major ways from Morris's study. First, the study was done with personal interviews, and not a survey. The participants were chosen with highly selective criteria; I contacted only those people who I knew had been involved in several collaborative endeavors. Choosing this method limited the number of participants to interview, but provided a rich set of

³Details of this study were presented in (Shah, 2010).

data for analysis. Conducting interviews also made it possible to customize some of the questions to the interviewee, which resulted in more meaningful responses. Second, in contrast to Morris's study, this study targeted collaboration in the information seeking domain in general, and not just for web search tasks. It is very unlikely that CIS is done as an independent process. Collaboratively seeking information is almost always done as a part of a larger project/problem, and it is important to study CIS in that context.

In the rest of the section, a synthesis is presented highlighting the lessons from these interviews. This synthesis is divided in three logical parts: instances (what), motivations (why), and methods (how).

3.2.1 Instances of collaboration (*What*)

From the description of their past collaborations, three major trends of doing collaboration were identified by the respondents:

1. *Forced collaboration*

Often people are forced to work together by their supervisor. A class project, where the instructor forms the groups, is such an example. The interviewees also mentioned working in such *forced collaboration* in case of a merger (permanent or project-driven). For instance, one of the participants, who worked on a welfare reform project, talked about how various agencies were brought together by a funding organization to work together on the project. These agencies may not have worked in collaboration otherwise, but as a requirement for their joint funding circumstance, they had to work together.

2. *Peer-to-peer situational collaboration*

This kind of collaboration was found most commonly among the participants.

The most typical example was co-authoring on research papers or proposals with colleagues. This was not surprising as the participants for this study work in fields that are highly collaborative, and most projects involve collaboration. Had participants from other fields such as humanities been interviewed, we may have found such collaboration less common. In addition to collaborating with peers on a professional basis, several of the participants also reported collaborating with their spouses for different reasons, with travel and shopping being the most common.

3. *Expert-novice asymmetric role collaboration*

As reported earlier, all of the participants were expert searchers. In addition to this, most of them were also subject experts. Therefore, they receive many requests for information on various topics that they are expert/familiar with. Faculty members obviously get constant requests from their students on their respective subjects, but even the graduate students are prone to such calls. As one of the participants, who was a graduate student, said about this participant's friends and colleagues requesting information from the participant, "They think I'm a walking library!"

Such requests often start collaboration. One of the participants described personal experience with responding to information requests regarding information architecture, on which the participant was considered an expert. The participant reported on some collaborations that were started by the participant in the past due to back-and-forth interactions with even some strangers based on those initial requests that the participant received. "Sometimes these requests turn into interactions, and into a relationship, and then we both start benefitting."

When asked about the group size for the collaboration, most reported working in a group of two to five people. Very rarely, the participants worked on a project that

involved eight to ten or more people.

The duration of a collaborative project varied a lot for each participant. The shortest duration of a collaborative project was reported as about two weeks, and the longest about a couple of years. A few weeks to a few months seem to be most usual duration for collaborative projects for the participants.

When asked for the optimal group size and project duration, everyone said that it depended on the situation. As one respondent said, “No matter how many people you put to it, pregnancy is a nine month affair!”

3.2.2 Motivations for collaboration (*Why*)

One of the most interesting questions while studying collaboration is why people collaborate. While collaboration is necessary in many situations, and beneficial in several cases, it has its costs and drawbacks. The interviewees talked about this issue from their personal experiences and observations.

Similar to the instances of collaboration, the motivations can be categorized as follows.

1. Requirement or setup

As reported by the respondents while talking about the instances of their past collaborations, often working together is a requirement of a project. As a respondent said, “Sometimes you don’t even think of doing a project any other way. You just have a group.” Merger of two organizations is such an example. While working on collaboration under such “motivation”, the respondents often found frustrations and loss of productivity. However, most participants pointed out that such frustration is primarily due to personal preferences. “Some people like to be told what to do”, as a faculty member testified referring to some of the participant’s students.

2. *Division of labor*

This motivation is one of the simplest appeals of many collaborations. For instance, one of the respondents (a faculty member) talked about letting the participant's students choose if they wanted to do an individual or group project, and most of the times the students chose to work in groups. This allowed them to distribute the work and get more done in the limited time and resources they had.

The distribution of work may not always be tangible, such as dividing up the searches to do or documents to read. As one respondent mentioned, sometimes one wants to use one's collaborators to cross-validate the work one has done. As another respondent pointed out, sometimes we also need to share the *load of thinking*. One respondent identified the value in having multiple people searching for information for a literature review. "Lit review is hard and the ACM Digital Library is bad," the participant stated.

3. *Diversity of skills*

This is the kind of motivation that makes it possible to have *the whole greater than the sum of all*. Almost all of the respondents admitted being involved in several collaborations because of this reason; either an individual did not possess the necessary skills to complete a task, or the participants realized the importance of involving people from different skill sets. For example, one of the respondents, a graduate student, discussed how the participant frequently collaborated with two other colleagues, one with statistics background and another with good writing expertise, the skills that the participant lacked.

A faculty member, very affirmatively, supported the idea of collaborating with people from different fields. The participant presented three conditions that the participant looks for while establishing such collaborations with someone: (1)

should like the person, (2) they should have fun working together, and (3) the process of collaboration should be stimulating and interesting. Another faculty member presented a point on the motivation for collaboration by saying that the participant expects to influence and be influenced. “..otherwise, what’s the point!”

Faculty members often involve their students in projects for educational purposes. One of the student respondents confirmed this and added that such collaboration with advisors help them learn new skills, and keep them focused. “[he/she] always knows the big picture,” the respondent said referring to the participant’s advisor.

3.2.3 Methods for collaboration (*How*)

The respondents identified several methods and tools that they have used or seen others using for collaboration. The most common among them were email, face-to-face meetings, IM, and phone or conference calls. In general, the choice of the method or tool for our respondents depended on their situation (co-located or remote), and objective (brainstorming or working on independent parts).

About half the respondents reported using Google Docs, but most of them were not satisfied with the results. One respondent complained about the difficulty in knowing ‘who did what’ on Google Docs, and the lack of support for using embedded spreadsheets.

A couple of respondents expressed their frustration with the Blackboard⁴ discussion board (one mentioned that it takes too many clicks), and reported success in using Sakai⁵ instead.⁶

⁴<http://www.blackboard.com/>

⁵<http://sakaiproject.org/>

⁶My personal experiences are not very different. The Blackboard discussion board was not very

Most respondents also reported successfully using whiteboards for collaboration, which of course, required them to be co-located and working synchronously. Using the ‘track changes’ feature of Word, on the other hand, got mixed opinions. Most respondents reported using it because there was no other or better alternative. This dissatisfaction sometimes drove them to simply make annotations on a printed document and pass it around in the group.

3.2.4 Summary

Following are some of the lessons and guidelines that were derived from the presented study.

- The so-called *entry points* for collaboration have changed significantly in the new millennium. The respondents who were active collaborators in a pre-web or pre-Google era talked about collaborations that used to begin only through explicit interactions and intentions among the participants. In recent years, though, due to wider accessibility of online tools, such as emails, web search services, IM, Skype, wiki, and blogs, people are able to have very low-cost interactions and loosely defined collaborations. The advent of Web 2.0 services have certainly sped this up.
- While the old philosophy of “birds of a feathers flock together” or the idea of homophily (Lazarsfeld and Merton, 1954) still prevails, collaborations that connect people of diverse backgrounds and/or span multiple domains are increasingly becoming common.

successful for my class and I had to switch to an other kind of online bulletin board, such as phpBB.

- Email and face-to-face meetings are some of the most popular methods of collaboration. These methods represent two extremes of the classical model of collaborative methods (Rodden, 1991; Twidale and Nichols, 1996), where email fits on the remote and asynchronous end, and meetings fit on the co-located and synchronous end. However, due to the changing structure of work environments and habits (people working on multiple projects with different sets of collaborators, across multiple sessions, and with multiple devices), the need to fill in the gap between these extremes is more apparent than ever. Figure 3.4 depicts various tools and methods on the classical model of describing the nature of collaboration. As we can see, systems and methods exist to fulfill most situations. However, transition from one situation to another may not be seamless, and that is where further work is needed. For instance, while talking about what they would like to see in a better system for collaboration, almost all the respondents desired to have a smooth flow between their different devices or environments with which they work. A typical example was being able to carry on a collaborative project on mobile devices and with or without the Internet connection.
- While most of the respondents wished for better tools for collaboration, they agreed they would have a hard time departing from familiar tools, such as email and IM, even though these tools were not explicitly designed to support collaboration. While tools such as del.icio.us exist, people still send website links to each other over email. System designers and developers face a grave implication due to this fact; they need to provide seamless integration of tools that support collaboration within a user's existing working environment rather than making him choose between his tried-and-tested method and a new tool. This finding reflects the views of Grudin (1994), where he suggested extending an existing single-user system, with which a user is already familiar, with groupware features

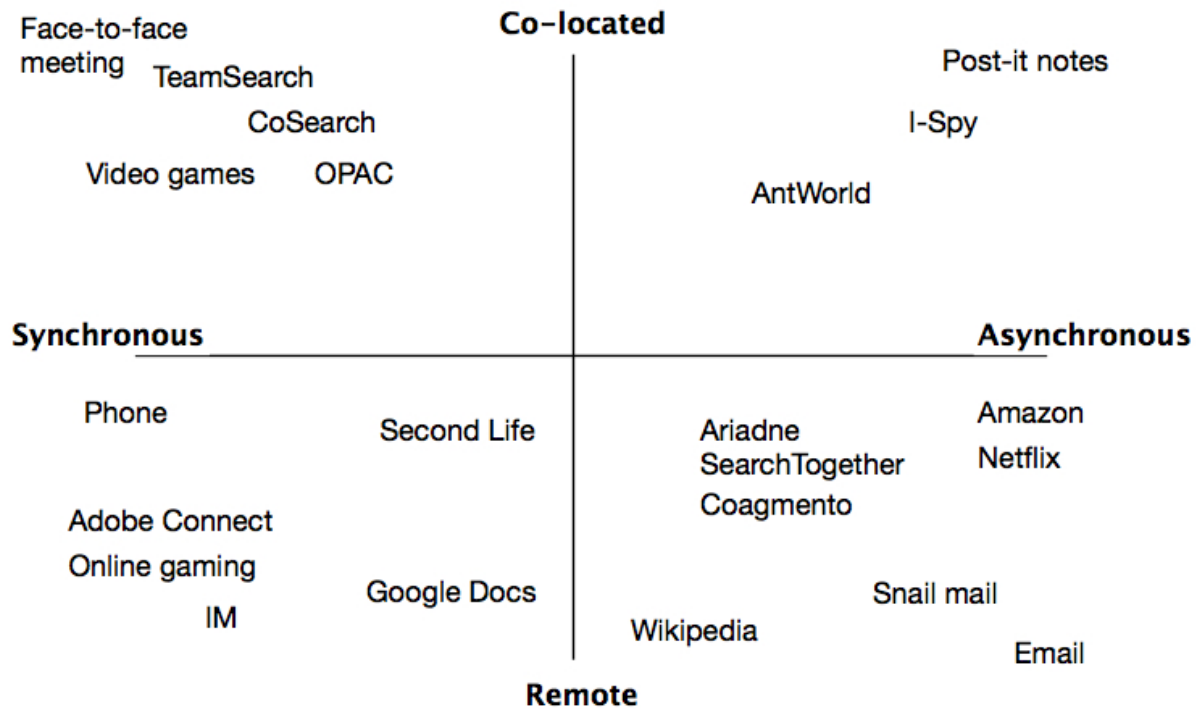


Figure 3.4: Collaborative systems/methods organized according to time and space aspects.

to minimize the cognitive load and maximize the adoption rate. As one of the respondent admitted, “We focus on results, and not how to do it.”

These points are highly valuable in designing a system that can support CIS. Based on the study presented here and the lessons learned, we can derive the following guidelines for designing a successful CIS system, as well as understanding its limitations.

- As we saw, often people may be forced to do collaboration. A CIS system, in such cases, can support that collaboration, but the success of such an endeavor depends on the ability of the participants in establishing the trust required to carry out the collaboration. If the participants really did not want to work together, the system would have no control over that. London (1995) and Gray (1989) have also presented such a realization in their works.

A CIS system should provide support for not only a well-established and intentional collaboration, but also for impromptu or forced collaboration with the caveat that if the participants really do not intend to collaborate or if they do not trust one another (or lose the trust in the process), the system cannot help them beyond making certain recommendations regarding the possible benefits of that collaboration.

In short, a good CIS system should be able to cater to the needs of (1) intentional and planned collaboration, (2) impromptu or forced collaboration, and (3) independent and individual work in collaboration. This leaves the act of collaboration up to the participants (if they have the intention or not), while having the support of the system whenever and however required.

- It was discovered during the interviews that people often use tools for doing CIS projects that were not designed for supporting collaboration. Morris (2008) also reported from a survey of knowledge workers that it is very common to send search results to a collaborator via email. As the author noted, people find their way around the traditional tools to make collaboration happen. Similarly, a CIS system that is designed to facilitate people's information seeking in collaborative projects should be able to support individual work as they would normally do with single-user systems. In fact, Surowiecki (2004) stressed that such an ability to be able to work independently is one of the requirements of a successful collaboration.
- Almost all of the interviewees reported collaborating to bring diverse skills to the table. Given this, it is important for a CIS system to highlight skills and actions of the participants in a collaborative project that can be useful to others in the team. For instance, a graphic designer may want to see her interface organized in a different way than a marketing analyst. If they are working on the same

project with a CIS system, we need to provide them with an ability to configure the interface in a way that suits their needs.

In an information seeking situation, this can mean saving and sharing a history of browsing, searching, and annotating. The system can take various documents found by different individuals and bring them together in a single pile of results. If there is a conflict, the system can notify the participants, so that they can discuss it and resolve the conflict.

A CIS system can even go a step further and use this information to make appropriate recommendations to the group members.

- While the respondents reported using traditional tools such as email much more than many specific tools for collaborations, they were open to the idea of accepting new systems, given such systems easily integrate into their existing practices. Fidel et al. (2000b) also recommended that, instead of imposing a rigid structure of a collaborative system to the users, one needs to configure their system around the way the users work. As Grudin (1994) suggested, one needs to develop a system that seamlessly integrates into the user's tried-and-tested methods. For a CIS system, this means providing interfaces for communication and sharing that are not only familiar, but also can be incorporated into existing tools on a user's working environment. If the users are familiar and happy with their email system, the new CIS system could try to provide the support for that email software right from its interface.

Another implication of this finding is that a CIS system should be easy to learn. It is expected that such a system will have its complexity, but to allow a new user to adapt to such a complex system, the system should be designed such that a novice user does not get overwhelmed in the beginning with all its features that

he may not find useful.

The participants in this study were also given a demonstrative walkthrough of Coagmento at the end of their interview session, to elicit feedback on the current version of Coagmento. The details of this process, along with the lessons learned, are provided in the following section.

3.3 Study-2: Demonstrative walkthroughs

This study was done in combination with Study-1. The participants were first interviewed and then given a demonstrative walkthrough. Thus, the set of participants for this study included 11 subjects from age 25 to 58 and with diverse backgrounds. These subjects were graduate students and faculty members in the fields of information science, library science, social science, and journalism. In general, these participants were mature researchers and/or academicians in their respective fields. A majority of these participants also taught various courses relating to information and library science, and media and journalism. They, therefore, could talk not only from their personal experiences, but also from those of their students.

I walked the participants through Coagmento's workflow in the following sequence.

1. First, the participants were given a brief introduction of how Coagmento is designed for serving two or more people working together in an information seeking task.
2. They were then shown Coagmento's main screen (Figure 3.22). Various components of the display were explained.
3. A sample search was run by me. When the results appeared on the screen, it was pointed out how Coagmento indicates documents that have already been viewed

by any member of the team, as well as any other attributes about a document in the result set, such as that it is saved or under discussion.

4. A document in the result set was clicked by me and, when the actual document was displayed in the browser, the participant was shown how Coagmento's toolbar (Figure 3.23) can be used for saving that document, putting it in the discussion box, or collecting a snippet from it.
5. Possible uses and scenarios of the boxes on the right side of Coagmento screen were explained.

After giving this demonstrative walkthrough, I asked the participants questions regarding the usability and functionality of the system. The questions that guided the interviews follow.

Q1. Which features of this interface have you already used in some existing system?

Q2. Are there features that you haven't seen before? If so, which do you expect would be useful for doing collaborative searches? Describe one use of each of the features in collaborative searching. How do these features map to what you have previously mentioned about your group project experience and/or wish list.

Q3. If "no" to new features, ask participants to map the familiar features to what they previously mentioned about group project experience and/or wish list.

Q4. Now that you have seen the functionalities of this interface, what else would you like to have for carrying out effective collaborative searching?

Q5. Would you suggest changing something on this interface to

1. make it more usable
2. make it more functional (useful)?

If answered “yes” to any of the questions, ask how.

One of the first questions asked was about identifying the components of the interface that they had seen or not seen before. Not very surprisingly, no one had seen all of the components in the same place. Most people had not seen color coding of the documents (based on views) or the query history. Most of the participants were also not aware of a system where they could collect the snippets, save the documents that are useful or keep them for later discussion. These are also the features that these participants found most appealing. Almost everyone appreciated having all the components and saw the value of them even in the situations where one was not doing collaboration. As one participant said, *“This is so cool; I could use such a tool even if I’m not working with others.”* Such remarks provided the support for appropriation for Coagmento, i.e., the unintended situations in which it can be used. It was noted earlier in Section 3.2.4 that it is important to design a CIS system keeping in mind that often people working in collaborative projects may desire to work individually. Demonstrative walkthroughs of Coagmento confirmed this aspect of the system.

On the flip side, some participants felt the need to extend certain features. For instance, about half the participants reported that they would like to see more meta-data about the queries and the saved/discussion documents, especially time-stamps. Similarly, one participant asked for a way to track the progress of the collaborative group by means of a timeline of the events (query issued, document viewed or saved, etc.). Three of the participants suggested having a workspace such as Google Docs to consolidate viewed, discussed, and collected information.

Most users felt comfortable having all the actions and results stored on the system server, but some wanted a way to export all that information so that they could use it in other applications. A couple of participants desired to have more configurable screen

layouts and customizable components.

The findings from this study can be summarized as follows.

- Ability to effectively collect and share information is highly desired from a CIS system. People typically do this using traditional tried-and-tested tools such as email and IM, without realizing that there could be more effective ways of doing it in collaborative projects. Coagmento allows the users to easily collect and share snippets of information, which most participants reported to be a unique feature. However, one may not want to use a CIS system such as Coagmento just because it has a couple of unique features. As noted before, we need to understand the larger context and environment in which people collaborate for information seeking projects.
- It was a pleasant discovery that the participants saw Coagmento's use even in non-collaborative projects. Such appropriation is important to have for a successful groupware system, as Grudin (1994) noted.
- The participants asked to see more information on the interface, including time stamps. This was important feedback, yet at the same time created a design challenge. How much information is enough and how much is too much? Should we show day, date, as well as time in hour, minutes, and seconds? We need to develop systems that allow the users to choose this. Having a configurable interface can let the users work through its features without getting overwhelmed.

Based on the lessons learned from the demonstrative walkthroughs, Coagmento was enhanced primarily to support a real life testing of the system. The TREC ciQA (Complex Interactive QA) 2007 data-set was used as the collection for this version of Coagmento. This data-set had nearly one million documents from various news

sources. The collection was indexed using the Lemur Toolkit,⁷ and a modified Indri⁸ search service served the requests in the background.

3.4 Study-3: Pilot runs

To test the effectiveness of Coagmento in terms of providing appropriate functionalities for people working together while seeking information, three pilot runs were conducted during three different SILS undergraduate courses at UNC Chapel Hill. These runs are listed in Table 3.1 and described in the following subsections.

Table 3.1: Pilot runs with Coagmento.

Run#	Date	Participants	Pairs
1	03/24/2008	12	6
2	04/21/2008	8	4
3	10/23/2008	16	8
Total	—	36	18

3.4.1 Procedure

At the start of a run, usernames and passwords were handed out in a random order to the students. That way, they were not able to pick their partners, although they might already know each other. At first, a brief overview of Coagmento was presented. Then they were allowed to login to the system. Following is the outline of the entire experiment.

1. Each user was given a drill task:

“You are reporters at New York Times. You are working on a story that could

⁷<http://www.lemurproject.org/>

⁸<http://www.lemurproject.org/indri/>

show a possible link between President Bush and Bono, the U2 Rock Star. In order to do this, you are first investigating what common interests they both have. Maybe they are involved in the same project or cause, or they like the same baseball team! It's your job to find out as many such common interests as you can. Find relevant documents and collect the snippets that have the related information on this topic."

When the users first login, they were presented with this drill task. They were also given a printed copy of this task, so that they could refer to it during the task.

2. Once they read the task description on the screen, they could start the task and were taken to the main interface screen of Coagmento.
3. The participants were allowed to try this interface out with the given drill task for about 5 minutes. One of the first things they did during this time was introducing themselves to their partners and finding out who their partners were. They were encouraged to use every feature of the interface.
4. They were asked to click 'Done' and the system took them out of that task.
5. The users were now presented with a new task:

"You are detectives, specializing in antiquities and historical documents thefts. Your current assignment is to find the evidence for transport of stolen antiquities from Egypt to other countries. Since such evidence often appears when such antiquities are returned to Egypt from other countries, you should search and file news about these goods being returned. Find relevant documents and collect the snippets that have the related information on this topic."

This was the real task and once again they were given a written copy of the task.

6. They were asked to take about 10 minutes to do this task, and allowed about 12 minutes before asking them to declare this task ‘Done’.
7. The users filled in the end of task questionnaire and the exit questionnaire.

At the end of this exercise an open discussion about this experience and the system was conducted.

3.4.2 Data description

This subsection provides description about various forms of log data collected during the pilot runs. Each piece of data is analyzed considering an individual as the unit, as well as a pair as the unit. While describing individual data, the unique and overlapping objects are computed for that particular individual during the session. For example, overlapping queries for an individual show the number of queries that the same individual re-used. On the other hand, while describing pair data, the unique and overlapping objects are computed within that pair. In the same example, overlapping queries are the queries that were re-used by either of the members of that pair that at least one of them had used before.

The participants (36 over three runs) used a total of 281 queries, with 163 unique queries. The distributions of number of queries for individuals and groups are shown in Figures 3.5 and 3.6. Both of these distributions are summarized in Figure 3.7.

Based on these figures, we can see that there is quite a bit of overlap in queries for individuals as well as groups. A very few individuals never repeated a query, but while considering the group as the unit, we can see that every unit had some extent of query re-usage.

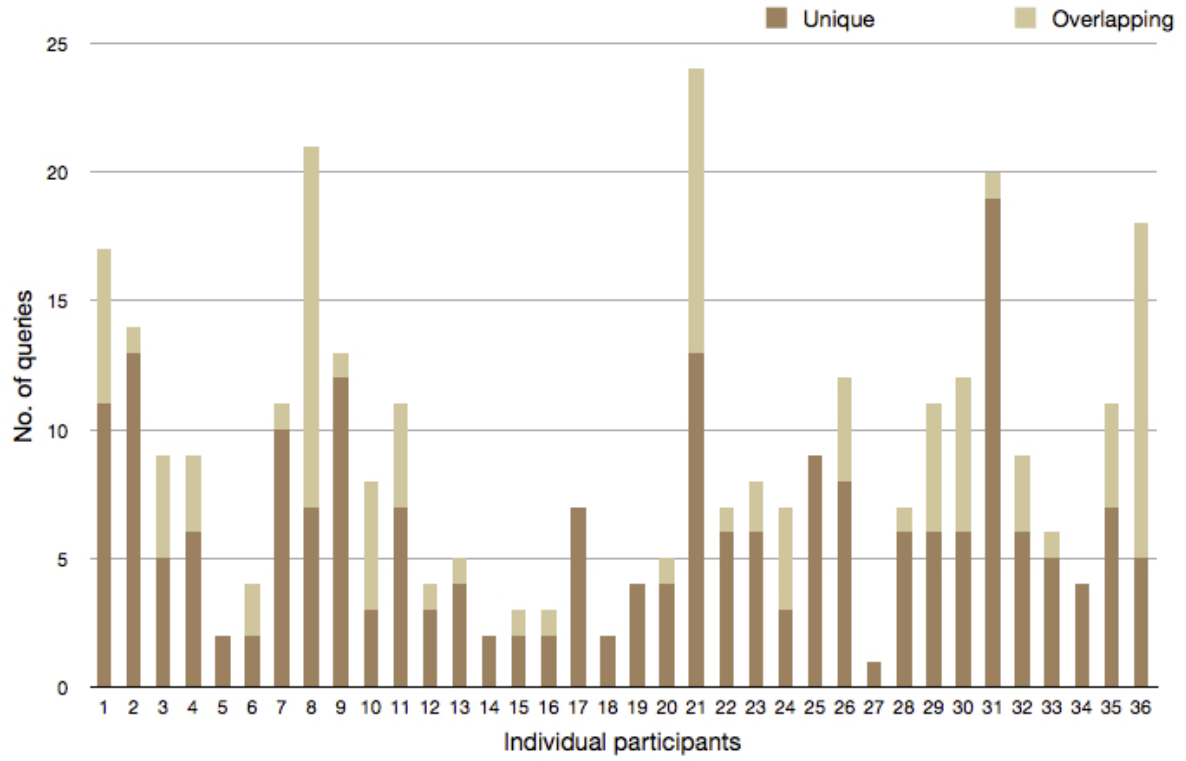


Figure 3.5: Query distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).

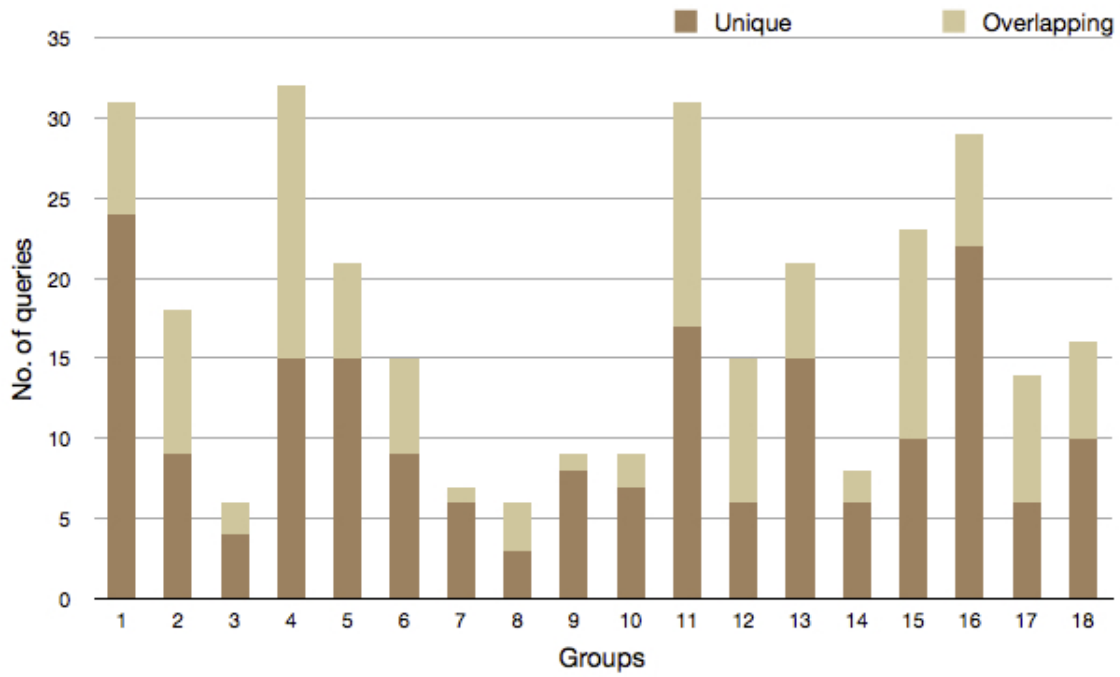


Figure 3.6: Query distribution for the groups.

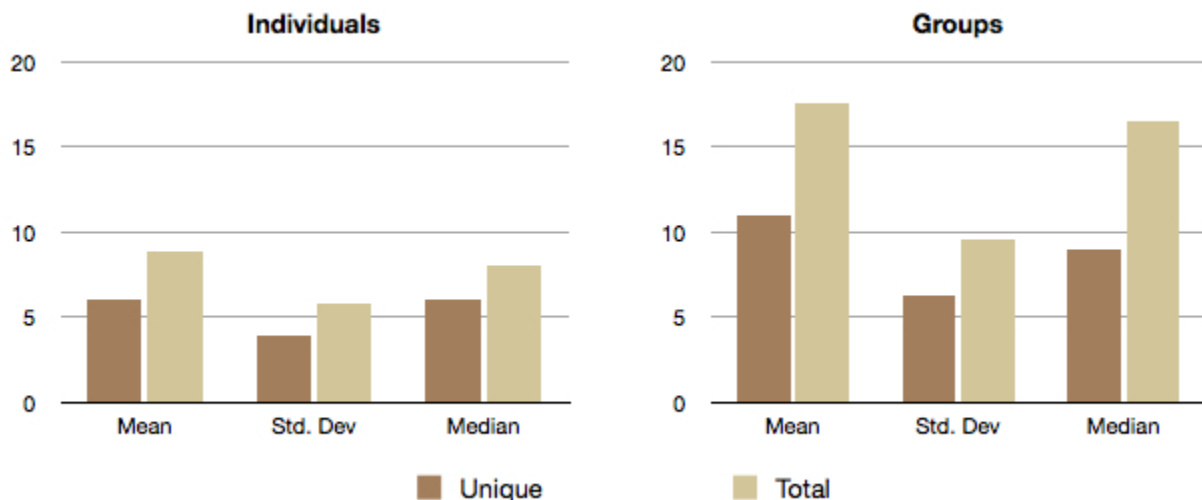


Figure 3.7: Query distribution summary.

The experimental version of Coagmento was designed to retrieve only up to 100 documents per query.⁹ It was found that over all the runs, the users saved a total of 318 documents, consisting of 89 unique documents. Figure 3.8 and 3.9 present the distributions of documents viewed by individuals as well as groups. These distributions are summarized in Figure 3.10.

As we can see, these distributions relating to the document viewing are very close to the distributions of queries. We can say there was a high correlation between using a query and viewing a document. However, unlike queries, we find less re-usage when viewing documents. Several individuals did not have any overlapping documents, which means after visiting a document, they never went back to it. Looking at the distribution for the groups, we find two of the groups (1 and 17) never revisiting a single document.

From log mining and discussions with the participants, it became clear that the participants rarely looked at a document already viewed by their teammates. Given the nature of the system, the participants had to start with a query to get a list of

⁹This turned out to be a limitation that some of the participants did not like, even though most participants never moved beyond the first page of the results.

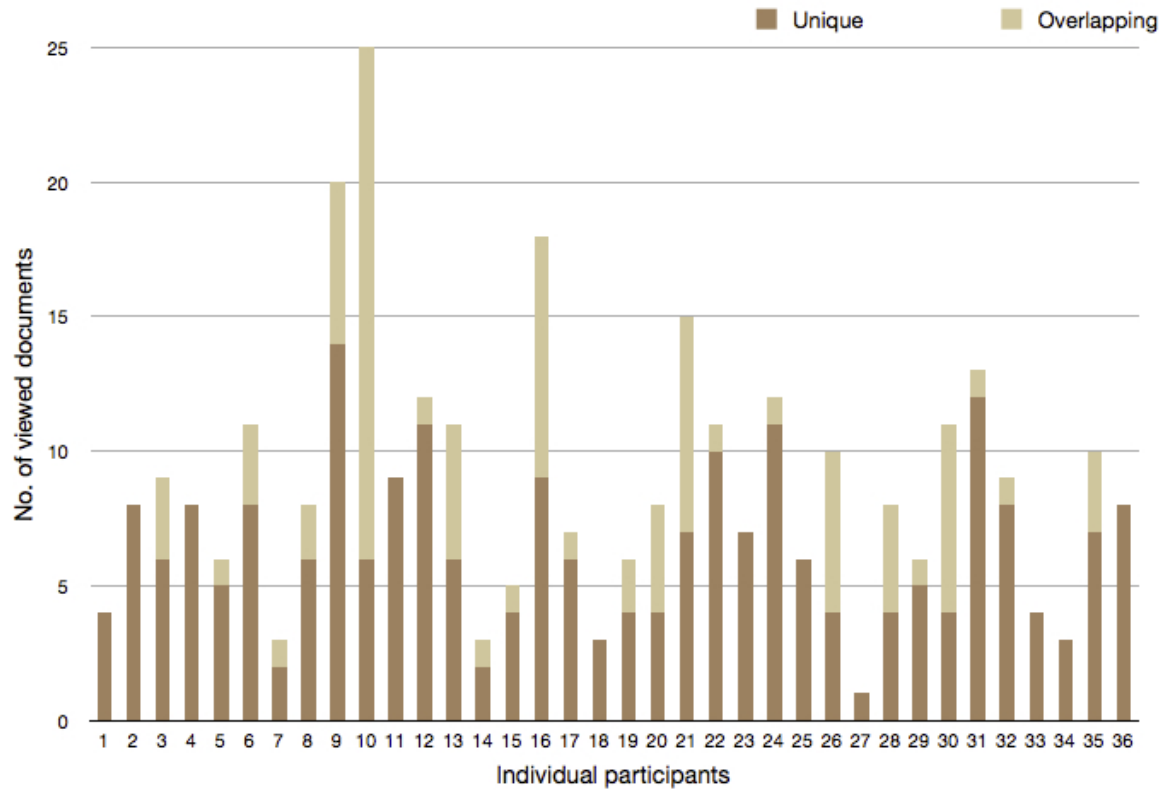


Figure 3.8: Viewed documents distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).

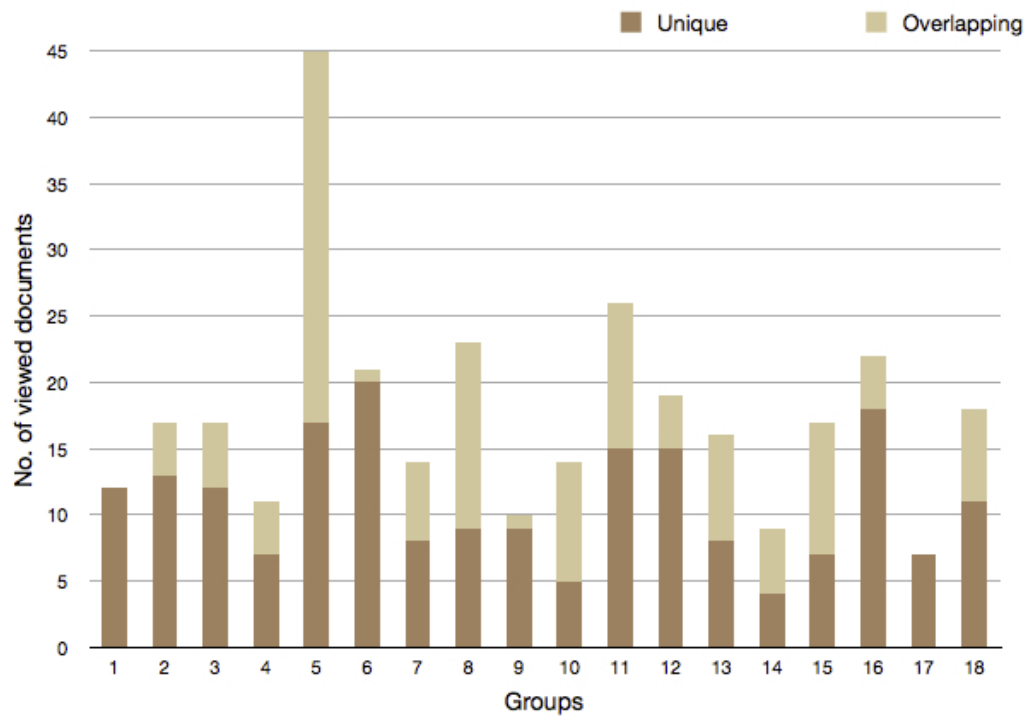


Figure 3.9: Viewed documents distribution for the groups.

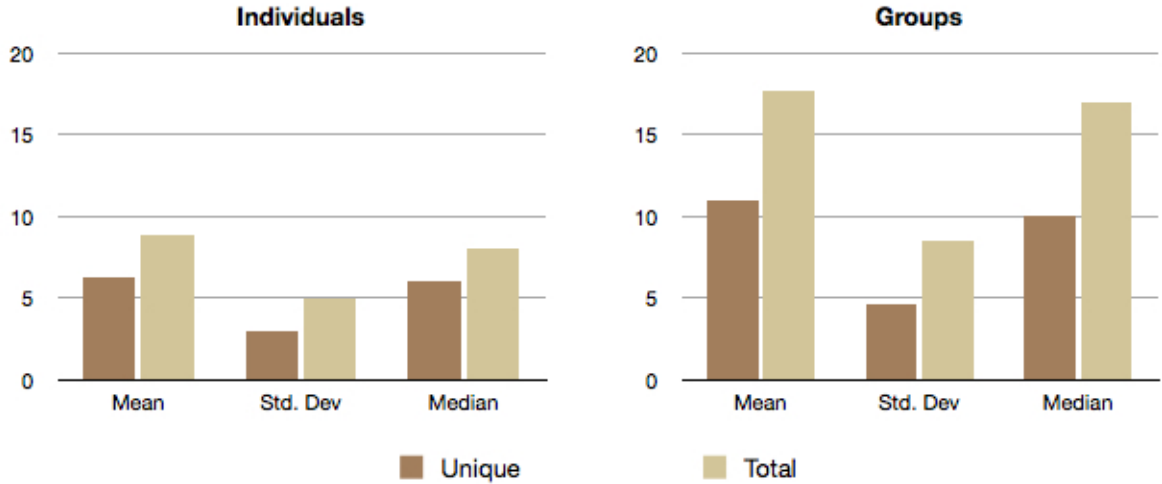


Figure 3.10: Viewed documents summary.

documents to look at. At the level of query formulation, the participants in the same team may use the same or similar queries, but once they get a list of results, they would avoid looking at each other's documents. This has two implications. First, if a task is time-bound, exploratory, and easily dividable, the participants may try not to do overlapping work. They may work individually trying to get as much information as possible, and then combine with their collaborators' individual information to create the group's product. Second, in order to easily know what has already been done by one's self and/or others in the group, the interface needs to provide ready support. This reaffirms the value of awareness in collaborative projects.

Of these viewed documents, the participants saved 99 documents, consisting of 44 unique documents overall. The distributions for the saved documents for individuals as well as groups are shown in Figures 3.11 and 3.12, and their summaries are given in Figure 3.13. From these distributions, we can see that, in most of the cases, the number of total documents was the same as the number of unique documents that were saved. This makes sense as, typically, one would want to save a document only once. For some reason, some of the participants saved the same document more than once.

It is possible that they accidentally pressed the ‘Save’ button multiple times for the same document.

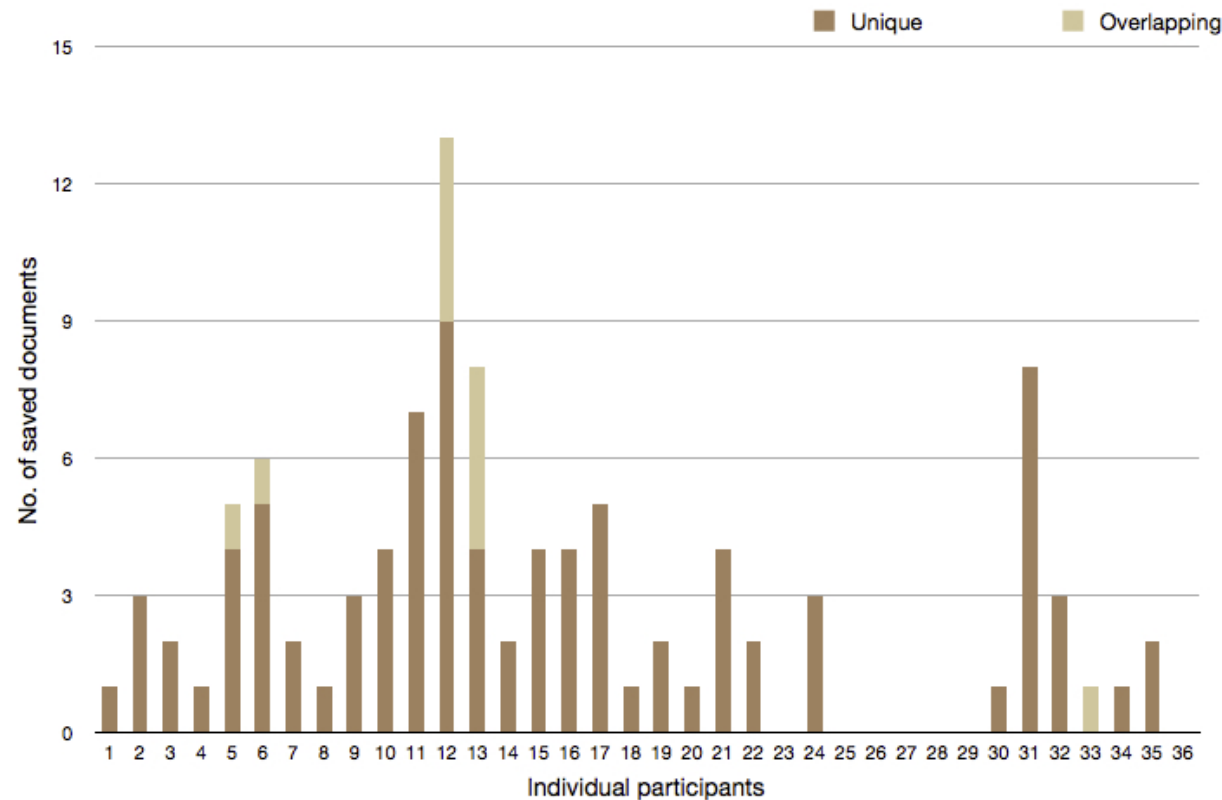


Figure 3.11: Saved documents distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).

We could also see some of the individuals affecting the group statistics. For instance, group 11 consisted of individual users 21 and 22. The group used a total of 31 queries, of which 24 were issued by user 21, almost 6 times as many as his teammate (user 22). Such disparity in individual performance within a given group can have significant impact on group productivity as well as an individual’s perception about the success of the collaboration. Further discussion of this topic will be presented in Section 3.4.3 using participants’ responses on post-task and post-session questionnaires.

We can also see that nearly half of the documents that were viewed were saved. This can indicate user perceived relevance, taken with the caveat that there was little

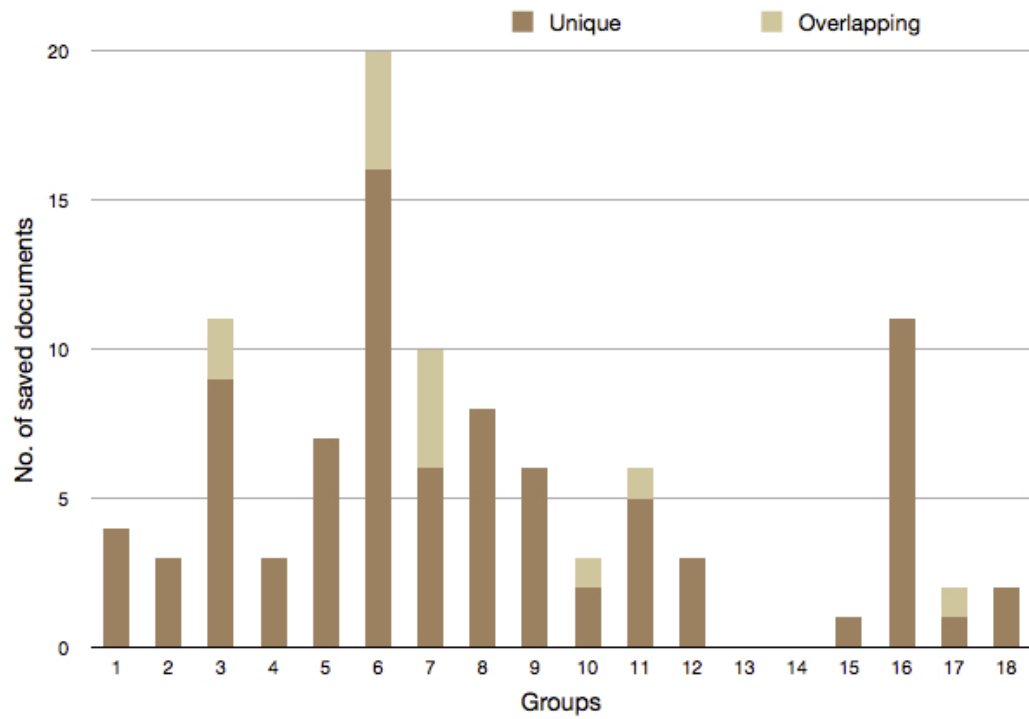


Figure 3.12: Saved documents distribution for the groups.

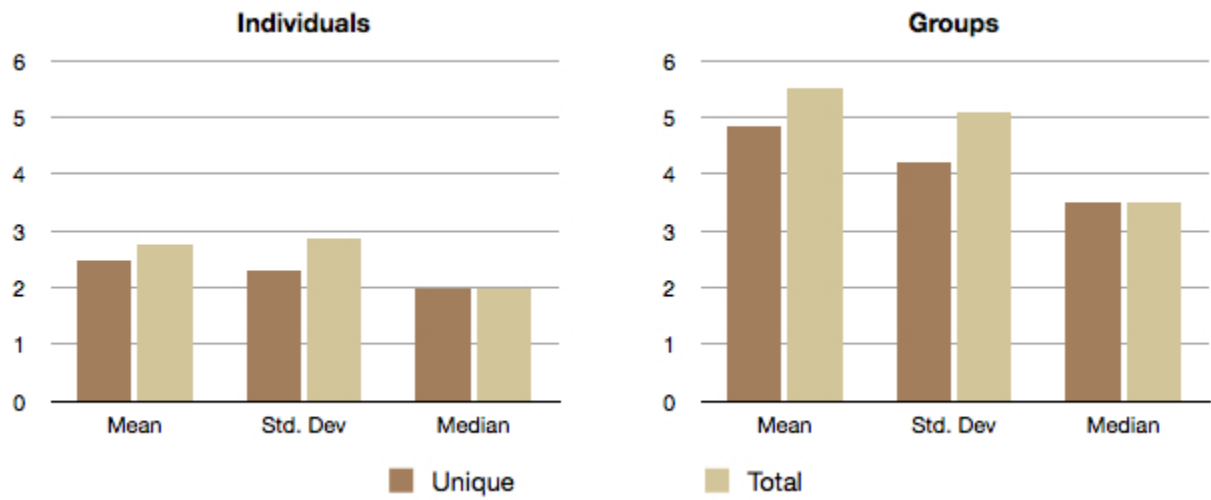


Figure 3.13: Saved documents summary.

cost to saving potentially relevant documents in a study where the participants did not have to do further processing.

The participants did not seem to be using the ‘keep for discussion’ feature much, where they can put a document for discussion/review with/by their partners. Overall, only 15 users used this feature and only 34 (24 unique) documents were kept for discussion, and none of them were transferred to ‘save’ or ‘discard’. Interestingly, ‘user08’ once transferred an already saved document to ‘discussion’, but that document was not discussed later and it stayed in the ‘discussion’ box. Such a behavior should not be surprising, given that the participants had only about 10 minutes to finish their tasks.

Coagmento allows the user to collect snippets by highlighting portions of a displayed document and clicking on the ‘Snip’ button on the interface’s toolbar. The distributions for collected snippets for individuals and groups are given in Figures 3.14 and 3.15, and their summaries are presented in Figure 3.16.

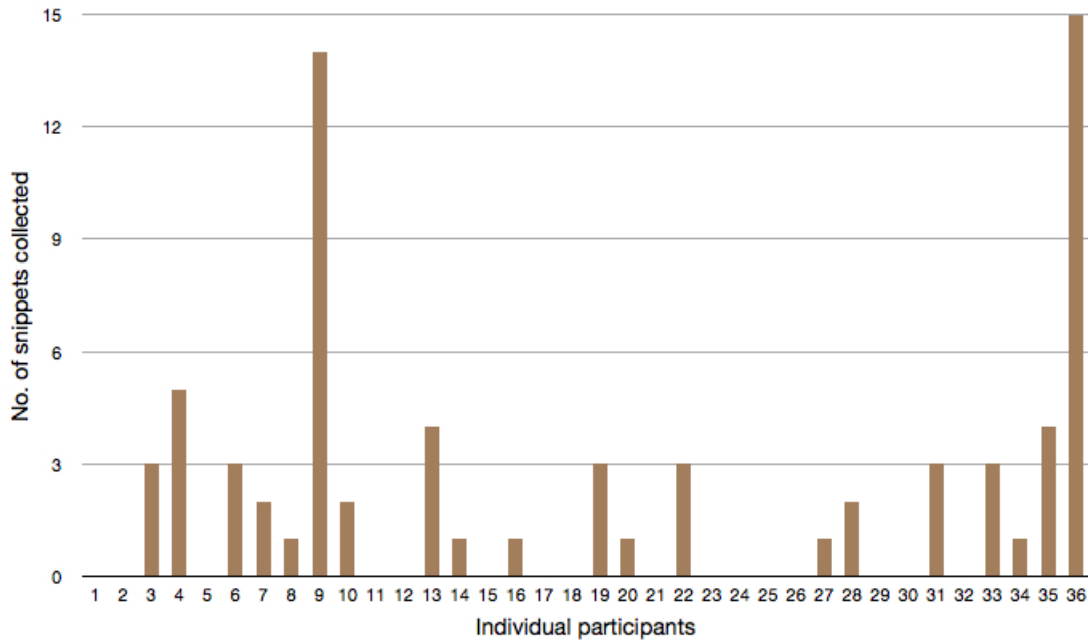


Figure 3.14: Snippets distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).

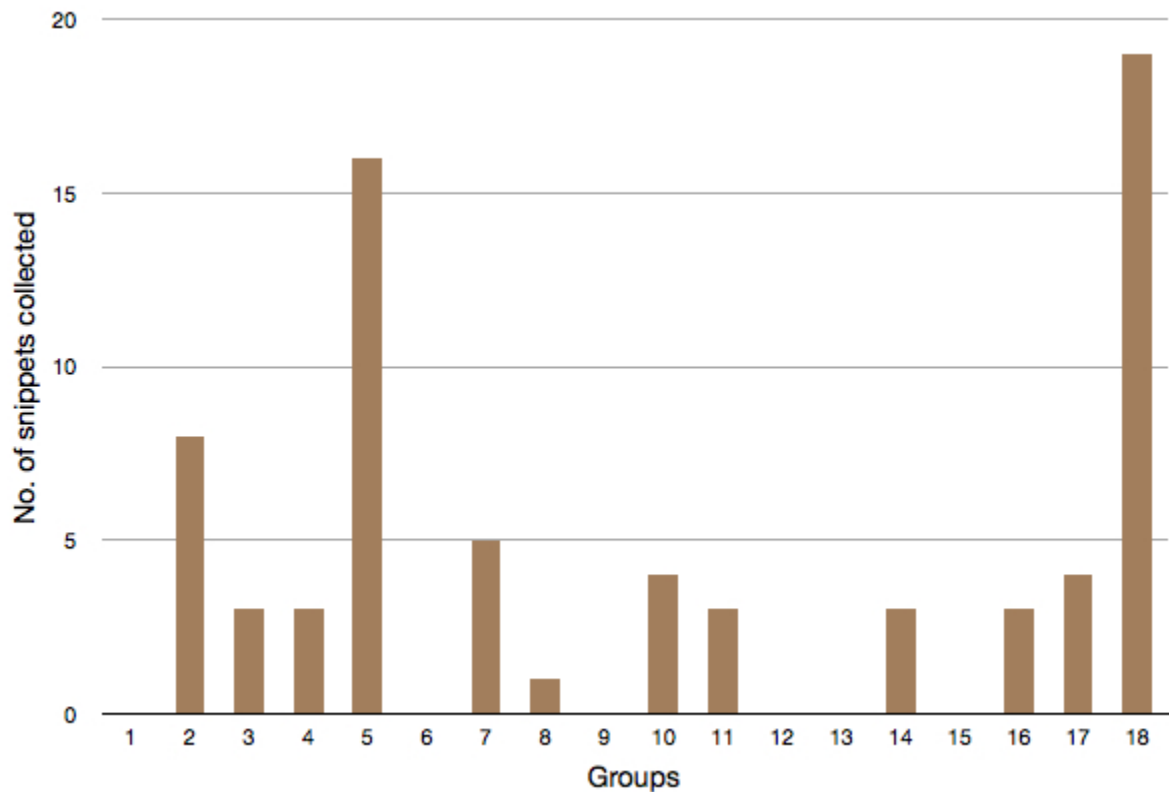


Figure 3.15: Snippets distribution for the groups.

While most of the participants liked this feature, not many snippets were collected. This could be a function of the task, the expectations from the participants, the time given, and the learning curve. It is highly likely that, if they were required to use the collected snippets later for preparing a report or doing a similar writing task, they may have collected more snippets.

Chat was the “trickiest” feature of this interface. This is mainly because of the fact that people are more used to chat/IM than any of the other tools provided. They had certain expectations while using this feature, which were not quite met. A total of 216 chat messages were sent between the partners. The distributions of the chat messages for individuals and groups are shown in Figures 3.17 and 3.18, with the summaries in Figure 3.19.

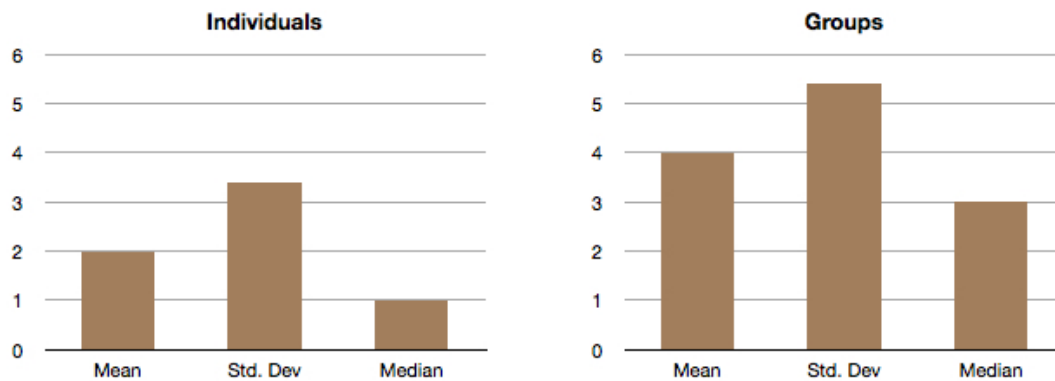


Figure 3.16: Snippets summary.

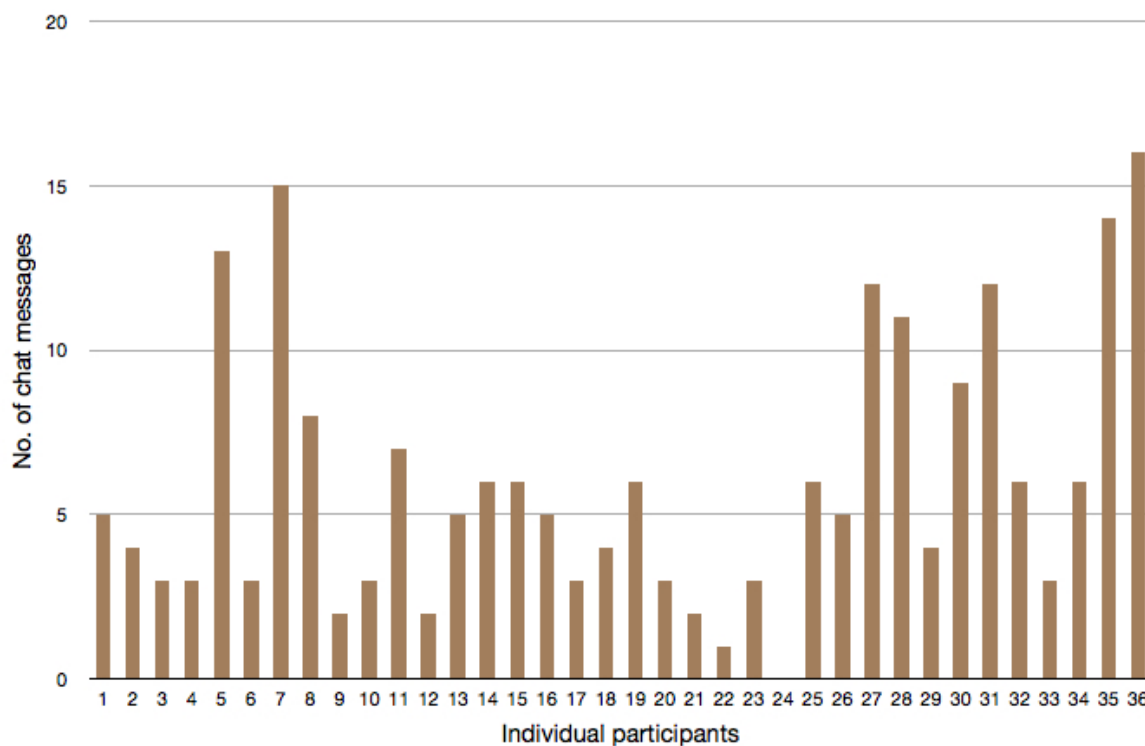


Figure 3.17: Chat messages distribution for the individuals. Consecutive numbers are pairs (e.g., 1-2, 3-4, etc.).

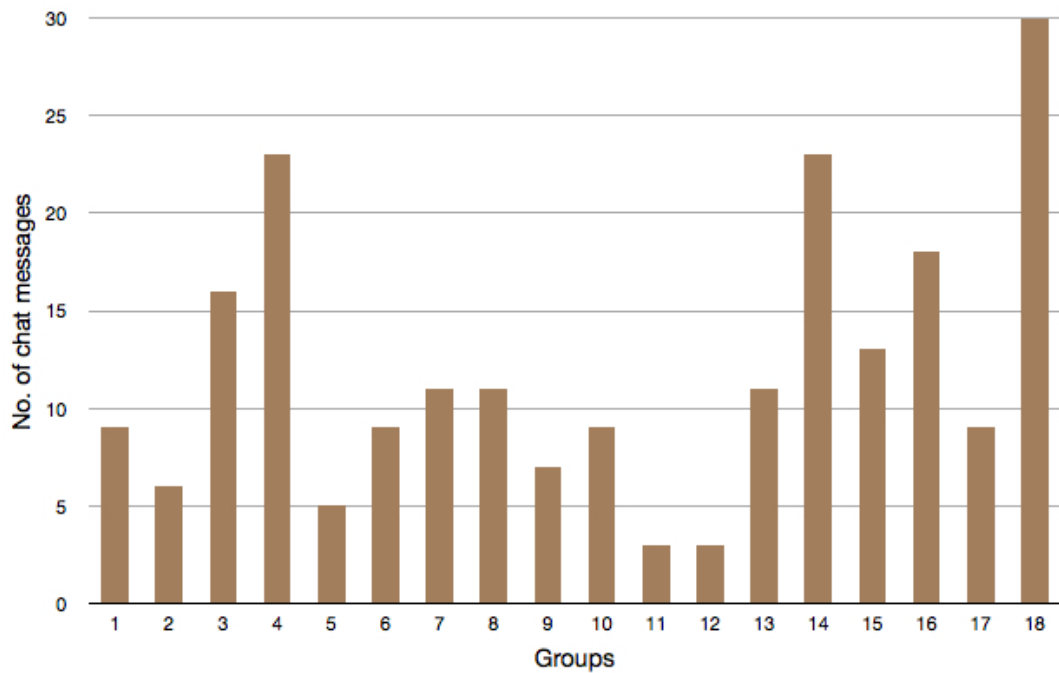


Figure 3.18: Chat messages distribution for the groups.

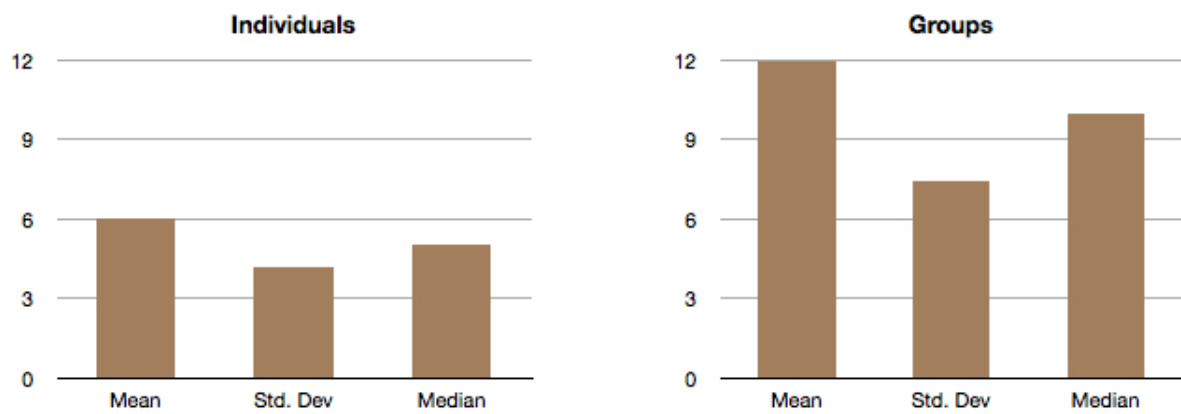


Figure 3.19: Chat messages summary.

Analyzing these chat messages, it became clear that most of the communications were not regarding discussion of the documents, but on more basic issues. For instance, one user asked his partner why he could get only 100 results. Many messages were simply funny remarks or casual conversations not directly related to task completion.

3.4.3 Analysis

The participants of these pilot runs were given two opportunities to express their opinions and evaluate the interface: at the end of a task, and at the end of the whole experiment. Since, for this pilot run only one task was given, these two forms of feedback were collected almost back-to-back. The questions presented at the end of finishing the task are given below. User responses to these are depicted in Figure 3.20.

Q1. Was the time enough to complete this task? (1=not enough at all, 5=more than enough)

Q2. State your confidence in your findings on scale of 1-5. (1=not confident, 5=very confident)

Q3. How satisfied were you with the search results? (1=not satisfied, 5=very satisfied)

Q4. How much did you feel the need to collaborate with your partner for this task? (1=not at all, 5=a lot)

Q5. How much do you think your partner contributed for this task? (Note: your partner will not know about this. So be honest!) (1=not at all, 5=a lot)

Overall, the users seemed to be well satisfied with this system, but that does not tell us much. Let us look at some particular cases.

‘user10’ ranked all the factors pretty low. Looking at his responses later, it seems

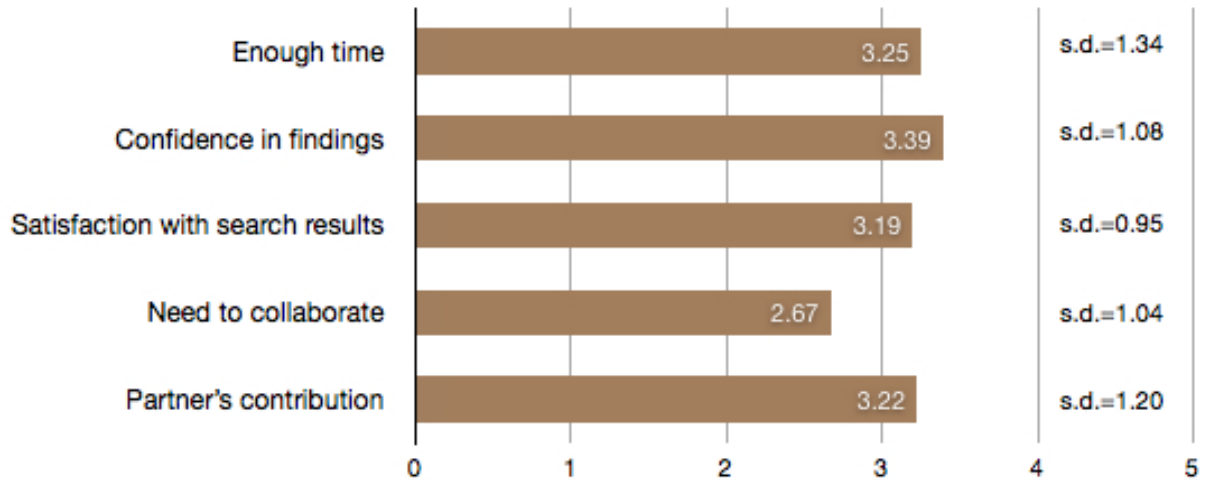


Figure 3.20: Post-task questionnaire responses.

that he was not very happy with the speed of the system and frustrated that it was difficult to get his partner's attention at times.

There were a couple of pairs in which both the users perceived asymmetry in their partners' contributions. For instance, in team-1, 'user01' did not think his partner contributed much, whereas his partner 'user02' thought 'user01' contributed a lot to this task. In team-6, both the users ('user11' and 'user12') were not quite happy with each others' contributions. In general, there was quite a bit of variation in the opinion of each user for this factor (min=1, max=5, median=3.5, mean=3.22, s.d.=1.20).

Following are the questions presented to the user at the end of the study. The first five questions asked the user to rate a factor between 1 and 5 and the rest of the questions required the user to input some text. All the users responded to the ranking questions, but not all filled in the free-form text fields. Responses to the first five questions are summarized in Figure 3.21.

Q1. Ease of use of the search interface on the left side (query box, search results)
(1=very hard, 5=very easy)

Q2. Ease of use of the other tools on the right side (chat, query history, review, saved documents, saved snippets) (1=very hard, 5=very easy)

Q3. Ease of use of communicating with the partner (1=very hard, 5=very easy)

Q4. Satisfaction with search results (1=not satisfied, 5=very satisfied)

Q5. Responsiveness of the system (1=not responsive, 5=very responsive)

Q6. Two things you liked about this interface

Q7. Two things you did not like about this interface

Q8. Any other comments

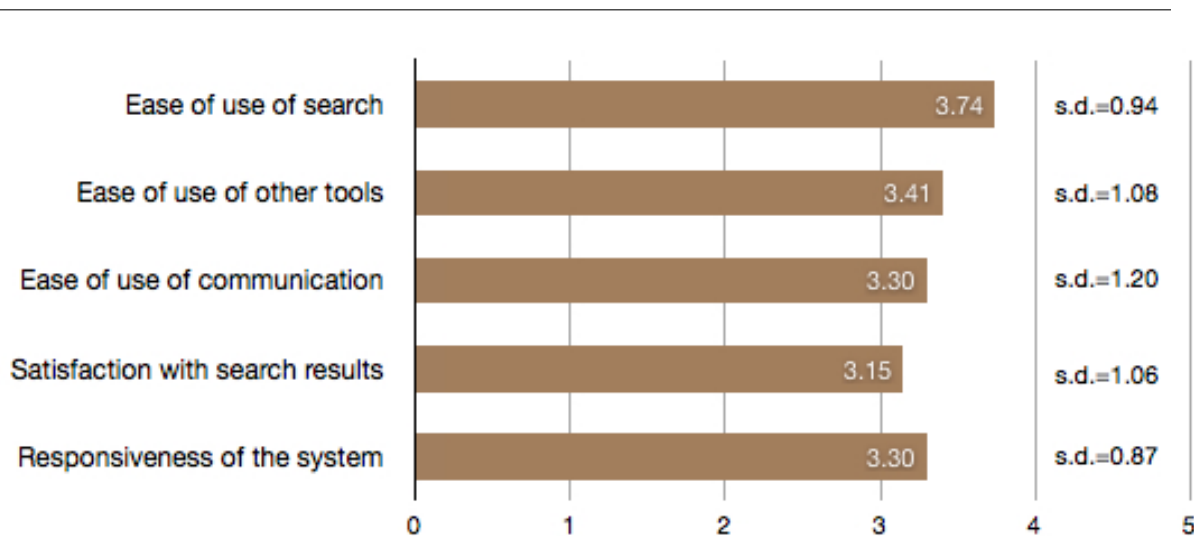


Figure 3.21: Exit questionnaire responses.

Once again, we can see that overall, users seem to be satisfied with the interface. On average, the responsiveness of the system was rated 3.30, which was highly correlated with the satisfaction with search results. It became clear during the class discussion that people tend to compare search systems with Google and anything that does not fetch them results as fast as Google does, is considered “too slow” or useless.

As a part of the exit questionnaire, the users were asked to provide free-form text feedback. This included questions about two features that they liked, two features that

they did not like, and any other comments.

It was found that more than half of the users thought chatting was an important feature. However, many users were unhappy with the way the chat feature was implemented.

Regarding viewing the documents, the users liked that they could immediately find out if a document was viewed or not and who had viewed it. A couple of people mentioned that they would have liked to see the difference between what actions they took versus what their partners did (saving a document, saving a snippet, etc.).

One of the most appealing features of Coagmento was being able to collect snippets. Most users found this very useful.

From additional comments that some users left, as well as the open discussion at the end of each run, it was inferred that people had two major issues with this interface: (1) the unrealistic situation, and (2) lack of familiarity with a non-Google-like interface. As one user mentioned, it was not clear to them why someone would use such a tool instead of a combination of Google, email, and IM. Several users also found the interface “not pretty”, “old style”, and “unorganized”.

By mining the logs, it was found that most users made use of the snippet feature. From the feedback that was received, it appeared that saving snippets was one of the most appealing features of this system. Users did not seem to use the ‘discuss’ feature much; they either did not put many documents in the discussion box or did not review the documents in that box often. This can be attributed to the short amount of time given to them for finishing the task.

3.5 Study-4: Participatory design

With Dr. Marchionini and Dr. Capra at UNC Chapel Hill, I conducted several sessions of participatory design in Dr. Marchionini’s HCI graduate seminar during Fall 2008

and Fall 2009 semesters. Each time, there were about a dozen participants involved.

The Fall 2008 sessions focused on discussing certain theoretical ideas of collaboration, and coming up with design specifications for an ideal CIS system. Since Coagmento was available to be tested by Fall 2009, those sessions focused on discussing various features and limitations of Coagmento and identifying ways to improve it. This section is logically divided in two subsections that discuss these two phases of the participatory design study.

3.5.1 Coming up with design specifications

We conducted four design sessions during Fall 2008 semester, that were several weeks apart. Some of the early discussions during these sessions were focused on definitions of collaboration, describing past experiences with collaborative projects, and identifying difficulties in such situations. Later the discussions were targeted to come up with design specifications for an ideal CIS system.

The participants mentioned several situations in which they were required to or wanted to do collaboration. Given that the participants were graduate students, a common situation was collaborating with co-authors on an article.

One of the issues that the participants reported during such collaborations is coordination. Often, the work is divided among the collaborators in the beginning of the project and a good amount of coordination is required to make sure that everyone is following a common timeline, and that their individual contributions come together in a meaningful product. For instance, in the case of co-authoring an article, the authors have to do their parts, exchange them with each other by a set deadline, and then arrange them to create the final write-up.

Most participants reported using general tools such as email, and specialized tools such as Google Docs for collecting and sharing information with their collaborators.

Some of the participants had used RefWorks for collaborative writing projects. While Google Docs was a common choice for writing-related projects, the participants identified several issues with it. They include not being able to attach files, not having search queries captured, and not having time stamp information stored with different actions.

Overall, the participatory design sessions in Fall 2008 helped reaffirm some of the issues that were discovered during the literature review and previous design studies, and helped in obtaining a variety of ideas from the participants through highly interactive group discussions.

Two major issues were identified during these sessions for designing a CIS system: (1) the role of awareness, and (2) support for a common workspace. The following suggestions were received for presenting *awareness* on the interface.

1. A tree or a map to see what path other collaborators took.
2. Personal histories for each of the persons in the group using a zoomable timeline or flowchart.
3. Provide history with different layers like Photoshop. One can then select the layer that he wants to look or work with, but finally they all come together to produce a finished picture.
4. Create switchable overlays of the interface, so one could switch between overlays depending on the requirement at the moment.
5. Provide a histogram of activities.
6. Have multiple tabs on the interface, like a modern browser, allowing one to have multiple views of the same project open and switch between them as needed.
7. Provide feed updates like Twitter.

The following suggestions were given with respect to the *workspace* that the participants would like to see in a CIS system.

1. Ability to choose one's role (e.g., supervisor, reader, information gatherer) and contribute through the workspace based on that role.
2. Sticky notes to remind one's self or give to other collaborators.
3. Use different attributes for/with different objects on the workspace to communicate effectively with each other. Such attributes include color, icons, size, location, gradients, avatars, and sounds.

A fair amount of discussion was also around the issue of searching in groups. The participants concurred that how one searches in collaboration depends on where the search falls in the whole continuum of creativity, and creativity is an individual thing. Given that, a system that lets multiple people search together and share their results, should allow the participants to express their individual creativity and then bridge multiple inputs to create a collaborative product.

3.5.2 Feedback on preliminary interface

The participatory design sessions for Fall 2009 were different in that, by this time Coagmento was developed and ready to be tested. This led to letting the students of Dr. Marchionini's Fall 2009 HCI seminar class try out an earlier version of Coagmento that supported multi-session projects, but not collaborations. The intention was to get feedback on various features of the system without involving collaboration.

This version of Coagmento was based on a new design architecture that is described in the next section. The participants were walked through its installation and given a brief tutorial on its usage. About a month later, they were asked to fill in an online

questionnaire rating various features of the system. These questions are given below and the participants' responses (on the scale 1 to 7, with each scale ordered from less to more) to them are summarized in Table 3.2. Only eight students responded, as providing feedback was optional.

Q1. It was easy to save relevant information (documents and snippets) using the toolbar functions.

Q2. Making annotations on webpages was useful.

Q3. Display of the project name in the toolbar was useful.

Q4. Display of various statistics about a displayed webpage (view count, snippets, and annotations) in the toolbar was useful.

Q5. Display of my personal history (queries used, and documents and snippets saved) in the sidebar was useful.

Q6. Ability to write notes using the sidebar was useful.

Q7. Writing a note using the sidebar was easy.

Q8. Log information about my activities (visited and saved pages, snippets, and annotations records) was useful.

Q9. Creating a new project was easy.

Q10. It was easy to learn to use this system.

Q11. I believe I became productive quickly using this system.

Q12. It is easy to find the information (logs and other objects, histories, etc.) I need.

Q13. The organization of information on the system screens (toolbar, sidebar) is clear.

Q14. This system has all the functions and capabilities I expect it to have.

Q15. I am able to efficiently complete my work using this system.

Q16. Overall, I am satisfied with how easy it is to use this system.

Q17. Overall, I am satisfied with this system.

Table 3.2: Feedback on Coagmento (version without collaboration support), averaged over eight responses.

Question #	Mean	s.d.
1	5.50	0.75
2	4.38	1.50
3	5.25	1.38
4	4.50	0.75
5	5.25	0.71
6	4.75	2.18
7	4.50	2.13
8	5.00	0.75
9	6.00	0.53
10	5.00	1.60
11	4.13	0.99
12	4.50	1.30
13	3.88	0.99
14	3.63	0.91
15	4.25	0.46
16	4.50	0.92
17	4.75	1.16

As we can see, the participants found it relatively easy to learn the system, use it for its intended purpose, and appreciate the usefulness of its various components, especially those related to providing awareness. In the free-form field asking them about two aspects that they liked about Coagmento, a majority of the participants identified the ability to save (bookmark) any webpage and collect snippets from anywhere as the biggest positive points. One of the participants suggested incorporating a citation builder like Zotero¹⁰ with snippets collection. In general, the participants appreciated

¹⁰<http://www.zotero.org/>

having ready access to their history (saved pages, snippets, and queries).

The participants, however, expected several other features not offered by the version of Coagmento they tried. This came up in Q14 (Table 3.2) as well as in the responses to an open-ended question asking them to list a couple of aspects of Coagmento that they did not like. One of the common feature requests that Coagmento did not have was the ability to save snippets from PDF files. This was identified as a highly valuable function in academia, since many scholarly articles appear in PDF format. Another requested feature was the way to organize collected snippets and notes.

In the discussion session followed by online feedback, the following additional issues came up.

- The participants reported experiencing an additional overhead for using Coagmento. They found the installation and learning to use Coagmento to be easy, but incorporating it in existing practices a bit hard. Due to the overhead involved with different actions, starting with login to selecting a project, and finally revisiting the collected information, the participants found themselves not utilizing Coagmento much.
- As a feature, Coagmento allows one to explicitly turn on and off its ability to record one's browsing. While this feature was provided as a way to facilitate privacy, it ended up being an obstacle sometimes as the participants forgot to turn it on. The participants were divided on the issue of making the recording on or off by default.
- Another issue the participants reported facing was of remembering to label a new project. Some of the participants also indicated that it was difficult for them to remember the actual name they gave to a project several days back. In other words, project management was below par and unclear.

Overall, the participatory design sessions provided a platform to engage in interactive discussions with several experienced participants, identify some of the key issues in designing a CIS system, and obtain feedback on early versions of Coagmento. These explorations helped drive the development of a new version of Coagmento, which is described in the next section.

3.6 Improved Coagmento

Based on the findings and experiences with previous studies involving Coagmento, it was once again redesigned. The version of Coagmento used in each of the pilot studies (except for the participatory design study phase-2) was website based, which means, that all the participants of a collaborative project have access to the various services such as searching, saving and sharing the results, and chat. An advantage of this approach is that it gives good control to the researcher over what the users do on this website. However, the biggest disadvantage is that the user is not able use the web as he wishes. Besides, the user may not always want to search. This realization led to redesigning Coagmento with a very different approach, while still offering the same functionalities as before.

The new Coagmento was developed with a client-server architecture, where the client is implemented as a Firefox plug-in that helps multiple people working in collaboration to communicate, and search, share and organize information. The server component stores and provides all the objects and actions collected from the client. Due to this decoupling, Coagmento provides a flexible architecture that allows its users to be co-located or remote, working synchronously or asynchronously, and use different platforms. This version of Coagmento used several of the suggestions and lessons derived from the design studies, while leaving out others due to the limited scope of this dissertation. In particular, Coagmento was redesigned considering the aspects of

awareness, communication, and ease of use of sharing while online information seeking. Thus, the aspects involving sophisticated interfaces for browsing and synthesizing shared information were not considered while creating the new version of Coagmento.

A screenshot of this new Coagmento is given in Figure 3.22. As we can see, it includes a toolbar and a sidebar. The toolbar has several buttons that helps one collect information and be aware of the progress in a given collaboration. The toolbar has three major parts:

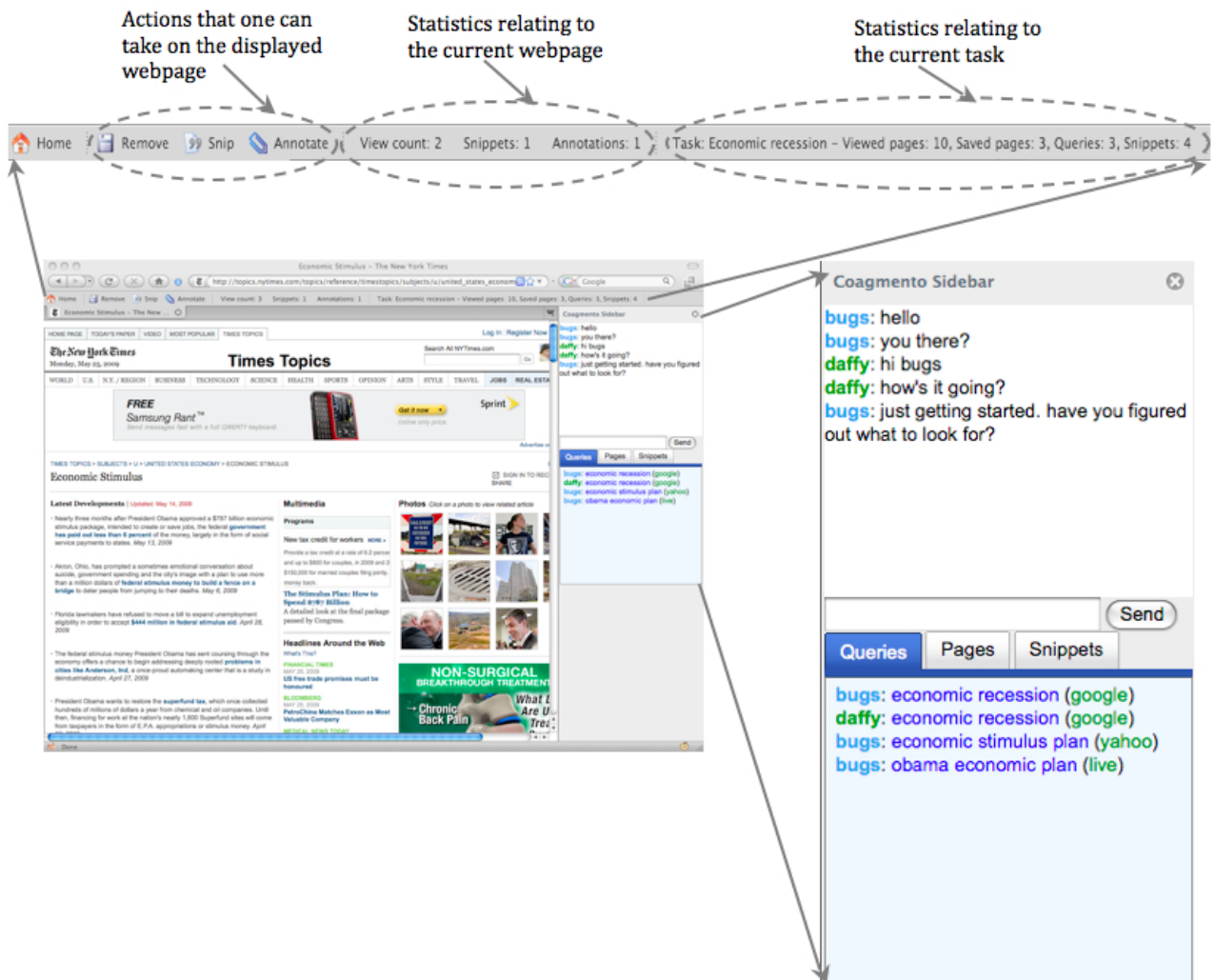


Figure 3.22: A screenshot of Coagmento with enhanced views of its toolbar and sidebar.

1. Buttons for collecting information and making annotations. These buttons help

one save or remove a webpage, make annotations on a webpage, and highlight and collect text snippets. The windows that pop-up while collecting a snippet and making an annotation from/on a webpage are shown in Figure 3.23 and 3.24.

2. *Page-specific statistics.* The middle portion of the toolbar shows various statistics, such as the number of views, annotations, and snippets, for the displayed page. A user can click on a given statistic and obtain more information. For instance, clicking on the number of snippets will bring up a window that shows all the snippets collected by the collaborators from the displayed page.
3. *Project-specific statistics.* The last portion of the toolbar displays task/project name and various statistics, including number of pages visited and saved, about the current project. Clicking on that portion brings up the workspace where one can view all the collected objects (pages and snippets) brought in by the collaborators for that project.

The sidebar features a chat window, under which there are three tabs with the history of search engine queries, saved pages and snippets. With each of these objects, the user who created or collected that object is shown. Anyone in the group can access an object by clicking on it. For instance, one can click on a query issued by anyone in the group to re-run that query and bring up the results in the main browser window.

3.7 Conclusion

The studies reported in this chapter helped shape the design and development of Co-agmento. This resulted in a better system, as well as an enhanced understanding of various challenges and issues for designing a CIS system. Some of the realizations from these studies are listed below.

Collecting a snippet from page: [Missed signals cleared the way for suspect to board plane - CNN.com](#)

Washington (CNN) -- To understand how the suspect in the botched terror attack was able to board a plane, you have to understand how the counterterrorism system that President Obama says failed is supposed to work.



Add notes to this snippet (optional)

This may come handy later.

Save Close

Figure 3.23: Snippets window.

Notes for page: [Cultivating Collaboration in the Cosmos](#)

 daffy	2009-12-31 13:42:46	This is the official blog of Coagmento.
 bugs	2009-12-31 13:44:04	I think we should keep coming back to this every week.

Save Close

Figure 3.24: Annotations window.

- It does not take long for people to learn such a new kind of interface. They may still not see the usefulness of all its features, but they can certainly start using them in a matter of a few minutes.
- There are certain tools that an average user is very familiar with and he expects certain kind of behavior from such a tool. In Coagmento, these were search results and chat box. For search results, people tend to compare with their familiar search services, such as Google. They want their results fast, and a lot in number, even if they never go past the first page. Similarly, almost everyone has used some kind of IM and they expect the chat panel to behave just like their favorite IM program.
- It is hard to make people collaborate in an unrealistic situation. Most of these users did not see the value in collaborating for such a common task as searching. The cognitive load for using the tools and/or collaborating was probably more than the potential return. However, it is quite possible that the same set of users would be more willing to use these tools the next time as they may have become more comfortable with them.
- It became clear that while most people saw the theoretical values of some of the tools, they were either not convinced that those tools were practical or refrained from using them extensively because of poor implementation. For instance, the chat feature was found useful by most users, but since the chat box in earlier versions of Coagmento behaved differently (it kept updating and scrolling up) than what they are used to, they did not use chatting extensively.
- People appreciate and desire unique features, such as collecting and sharing snippets of information effectively. However, such features themselves may not be compelling enough to adopt the whole system. One needs to understand these

features in a larger context and provide proper support for that context as well. In the case of snippets, it was discovered that, while effectively collecting snippets is a desired feature, the users want to be able to use those snippets in some way. This may involve organizing the snippets and using them to synthesize a report. Thus, without the support for this larger context of being able to utilize the snippets, allowing the users to simply collect the snippets may not be a strong reason for system adoption.

- Using a new CIS system is not simply about adopting a new tool, but it is about getting used to a new way of thinking for exploratory and collaborative work practices. Managing a multi-session collaborative project that already has enough complexity could become even more difficult and confusing while using a completely new interface, as reported by our participatory design participants. It is valuable to provide appropriate support for control, communication, and awareness; but, if it is done without understanding existing practices and cognitive load required to adopt the system, we may end up hurting its usage and adoption.

Chapter 4

Experiments and Results

The previous chapter reported the evolution of Coagmento, a CIS system. Having obtained various forms of feedback and gone through several revisions, the system was evaluated in a large scale study that explored some of the key issues in user-centric CIS projects, mediated by a specially designed system like Coagmento. This chapter provides details of a laboratory study that aims to investigate such issues, including the role of awareness, user satisfaction, and engagement in CIS projects that span multiple sessions.

Lauwers and Lantz (1990) argued that, for supporting user-centric transparent collaborations, we need to develop systems software that is distinctly collaboration-aware. They called this *collaboration awareness* and suggested that it is a necessity for shared and collaborative work environments. Coagmento is such a tool, providing specific kinds of collaboration as detailed in the previous chapter, and will be used for the experiments presented here.

As far as the design goes, several design decisions were already made for Coagmento based on previous works, demonstrative walkthroughs, pilot runs, and general understanding of creating a familiar, easy to use, and accessible interface. Therefore, the primary objective of the present work was not to study the optimal design or compare

different designs for a CIS system, but to evaluate the impact of different awareness components for their effectiveness, efficiency, usefulness, and usability in supporting collaborative activities and inducing higher levels of satisfaction and engagement in collaboration.

4.1 Background

The importance of studying awareness in collaborative projects, as well as methods for providing awareness were highlighted earlier in Section 2.6.3. Since the focus of this dissertation is on developing a framework for user-centric CIS, and the value of providing support for awareness in such situations has been substantiated using the literature review and preliminary design investigations, the primary study that is reported in this chapter will specifically address the issue of awareness in CIS systems.

The design of Coagmento (described in Section 3.6) was greatly influenced by SearchTogether, but unlike SearchTogether, Coagmento provides tools and support for not only searching and sharing, but also for organizing and synthesizing information. In addition to query and results awareness, Coagmento incorporates contextual and workspace awareness. This made it convenient to explore the impact of various kinds of awareness in a CIS environment.

To my knowledge, there has not been any recent study in the domain of online collaborative information seeking that looks at awareness in the context of different interface designs. The reported work here, thus, is highly valuable in adding to our understanding about awareness in CIS.

In the present design of Coagmento, various kinds of awareness discussed earlier are mapped to system components as follows.

1. *Group awareness* is facilitated by providing information about one's partner's current status (online/offline) and activities (e.g., document or rank list being

viewed).

2. *Workspace awareness* is given by providing a common space where the group members can discuss and compile their collective results.
3. *Contextual awareness* is instantiated by display of information about the task and the goals.
4. *Peripheral awareness* is given by presenting a history of various actions such as documents viewed and queries used, as well as products such as saved documents and snippets in a way that does not interrupt a user's current activity.

The study reported here varied conditions that included or excluded these awareness components. More specifically, the study manipulated peripheral awareness in three different ways by providing: (1) no peripheral awareness, (2) personal peripheral awareness, and (3) group peripheral awareness. The value of focusing on peripheral awareness is inspired by some of the works in the CSCW literature. For instance, Bly, Harrison, and Irwin (1993), with their experiments in media space, found that even though it was seemingly invisible, peripheral awareness was the most useful one. Gaver (1992) also argued for the provision of peripheral awareness that provides unobtrusive information of the people, artifacts, and environment, while working on a collaborative project.

4.2 Method

This section discusses a laboratory study conducted to understand the impact of various forms of awareness support in CIS tasks. The study involved several pairs of participants working on two different tasks over two separate sessions. The details of this study are given below.

4.2.1 Participants

84 participants in 42 pairs were recruited from UNC Chapel Hill. These participants were asked to come to the lab for two different sessions, which were one to two weeks apart. The participants were able to choose the day and time convenient to them using the online recruiting system (see Figure C.1 and C.2). Since the participants had to sign up in pairs, both the participants in a given pair already knew each other. In addition to this, it was required that the participants in a given pair should have done some collaborative work with each other before; thus, making sure they not only know each other, but also are comfortable working with each other on a collaborative project. The approval of a pair's participation in this study was based on these criteria. Participants were compensated \$25 each for their participation in two sessions. There was also a prize for the best performing team (two iPod Nanos). The performance was based on three criteria: amount of information collected, quality of this information, and teamwork.

Of the 84 participants, 27 were male and 57 were female, and their ages ranged from 17 to 50 with a median age of 21. Several of the pairs were co-workers or spouses. A majority of the participants were undergraduate or graduate students, while a few were university employees. The questions used for obtaining demographic information is given in Appendix D. Responses to some of the questions are consolidated in Figure 4.1. In addition to these questions, the participants reported their search experience to be 5.31 (s.d.=1.108) on the scale of 1 to 7. With the same scale, they also reported they enjoy working on collaborative projects (mean=4.63, s.d.=1.220) and have had good success on these projects (mean=5.74, s.d.=0.866).

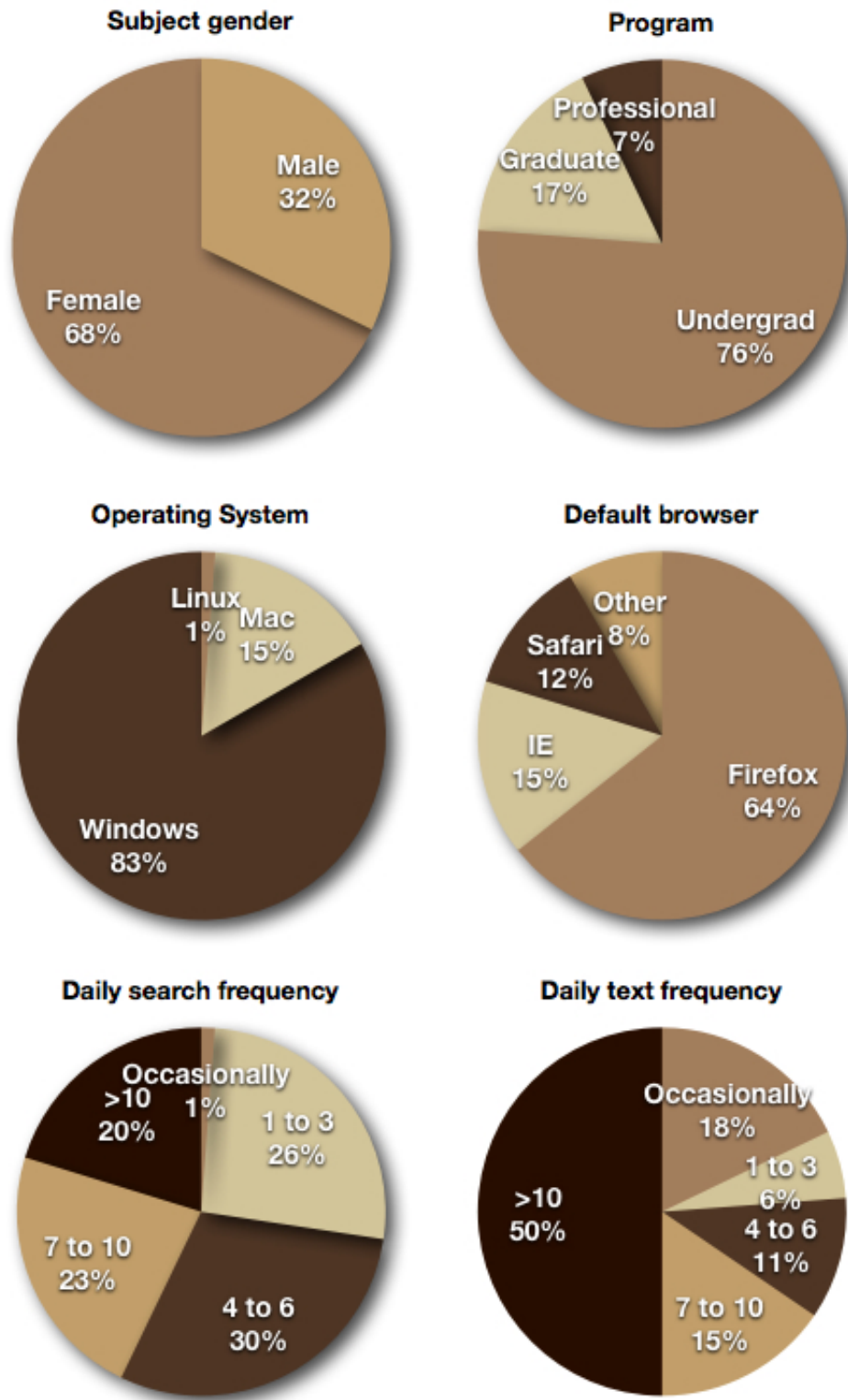


Figure 4.1: Responses to demographic questions.

4.2.2 Conditions

The participants were assigned to one of the three conditions randomly (14 per condition). These conditions were defined based on the provision of different levels of awareness-related support to the participants. They are as following.

1. *Baseline* (Figure 4.2): support of contextual awareness (current task name and topic description), and workspace awareness (where the team can see their collected objects - webpages and snippets - useful for the task).
2. *Personal peripheral awareness* (Figure 4.3): support of personal history (documents and snippets saved, queries used, etc.) in addition to the support provided in the baseline.
3. *Group peripheral awareness* (Figure 4.4): support of group history (documents and snippets saved, queries used, etc.) in addition to the support provided in the baseline.

Thus, the main independent variable here was the kind of peripheral awareness provided (1=none, 2=personal, 3=group). All the conditions provided support for communication, contextual awareness, and workspace awareness. No condition had the traditional kind of group awareness (where a participant can see exactly what another collaborator is doing at a given time). This was due to two main reasons. First, since the participants were required to work on the same project at the same time, they knew the other person was not only online, but also working on the same task. This realization eliminated the need to know the immediate status of one's collaborator. Second, introducing this additional condition would weaken the statistical power of the study. Therefore, it was decided to omit the traditional group awareness feature from this study interface. However, participants in each of the three conditions had the same support for communication, which could be used to check each other's specific status.

Actions that one can take of the displayed webpage

Name of the task

Chat

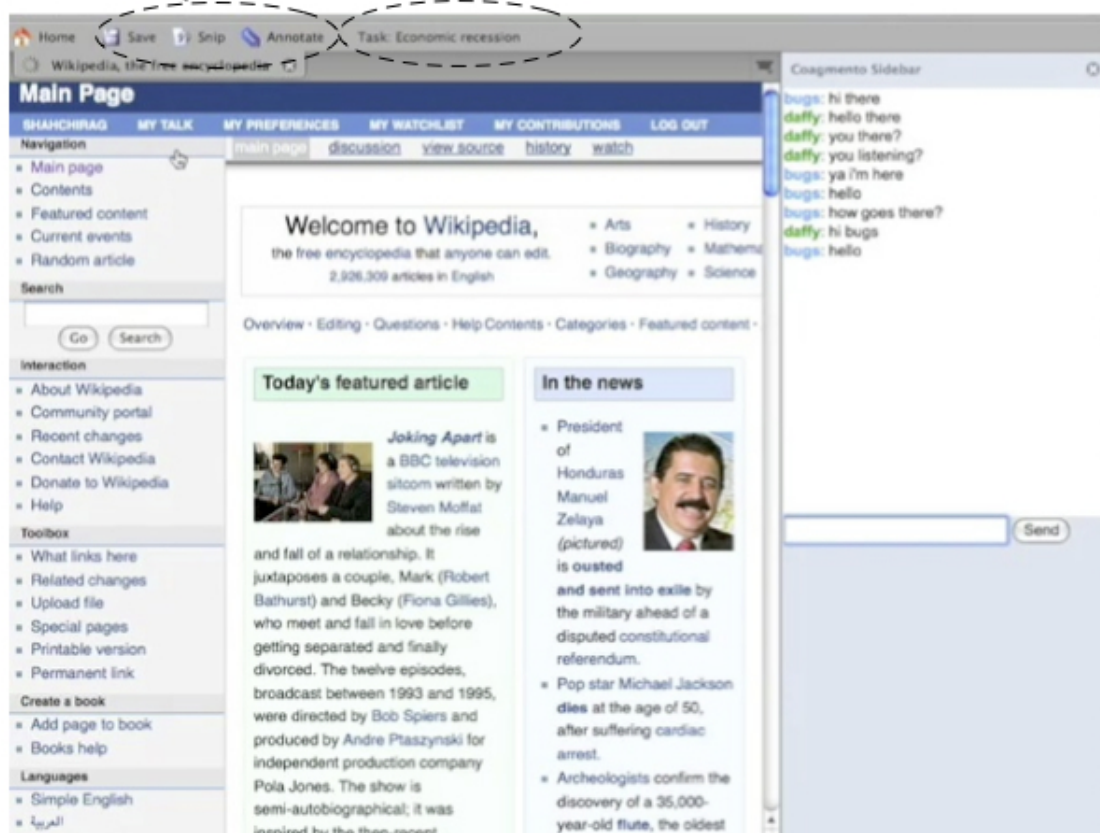


Figure 4.2: Interface for condition-1, baseline.

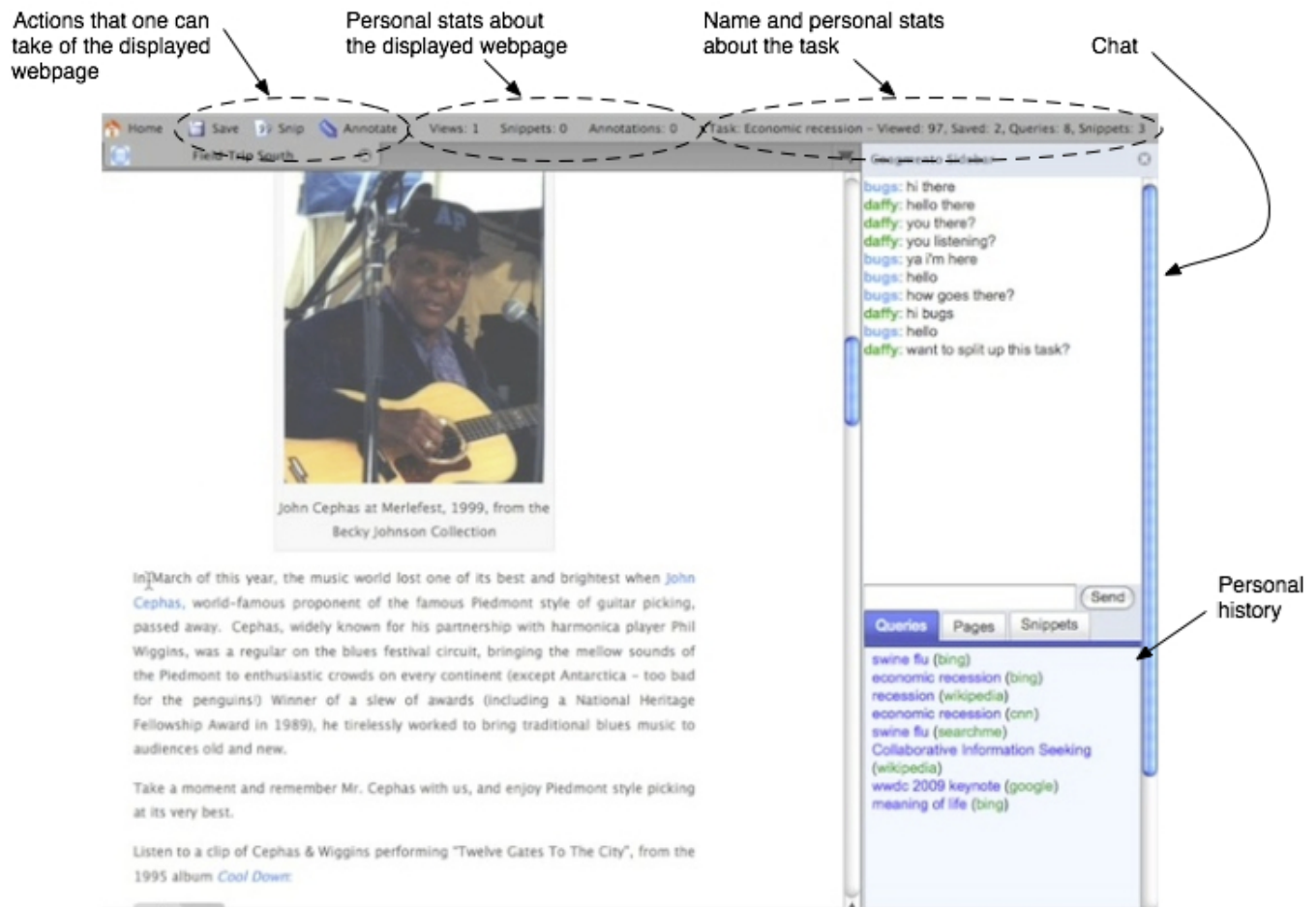


Figure 4.3: Interface for condition-2, personal peripheral awareness.

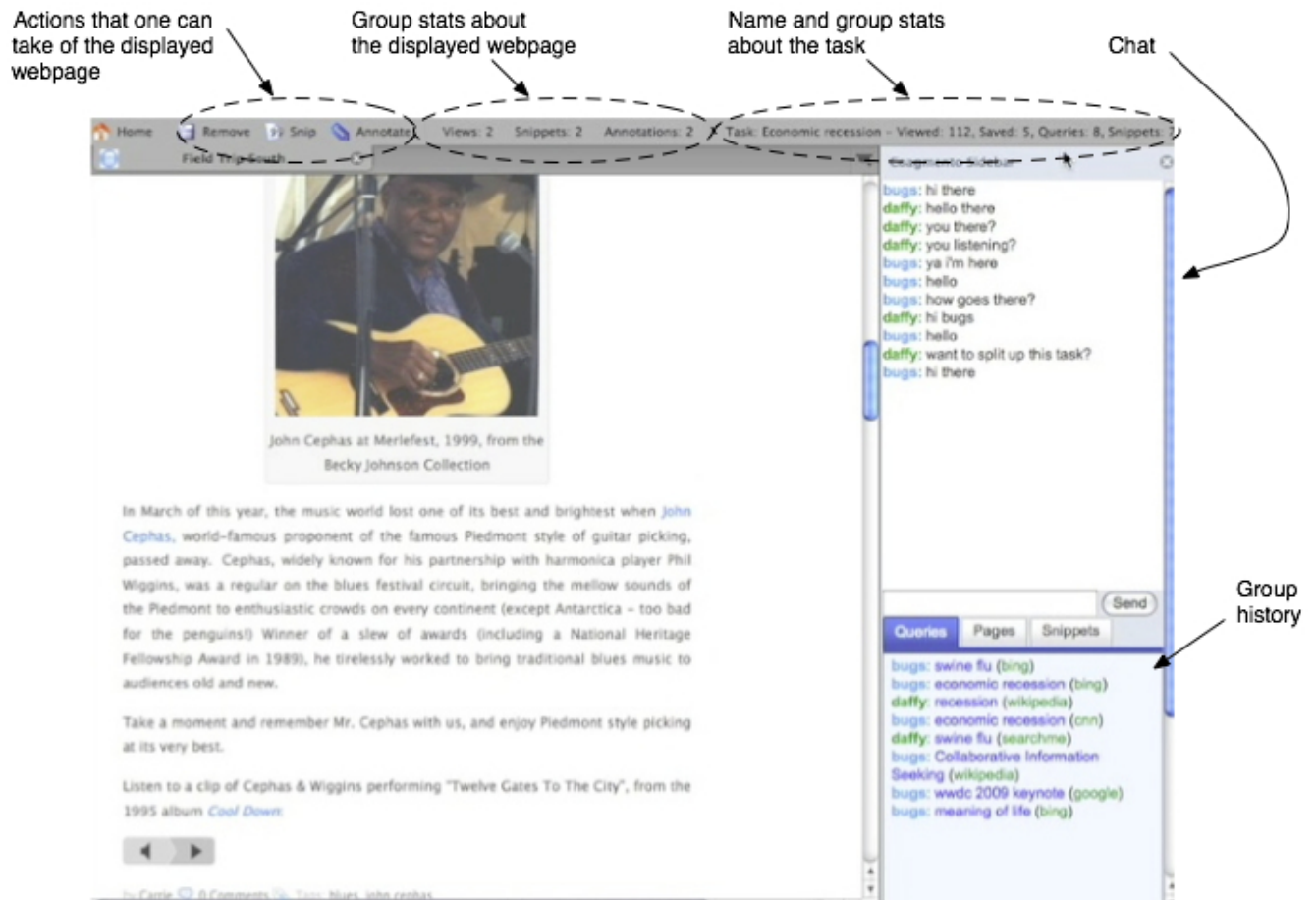


Figure 4.4: Interface for condition-3, group peripheral awareness.

Note that peripheral awareness does not indicate the nature of the information, but how it is presented. For the purpose of this study, it is divided into personal and group peripheral awareness, corresponding to the information relating to personal and group history, respectively. Both conditions 2 and 3 had the same components and presentations for instantiating peripheral awareness, but the information displayed there is different.

To support workspace awareness, each participant was given access to a common workspace. This workspace is updated in real time and shows the objects (webpages and snippets) collected during the information seeking activity (Figure 4.5), and switches to a different view showing collected snippets and a way to organize the snippets during the information compilation activity (Figure 4.6). With each object displayed in the workspace, author information (who collected that object), and corresponding notes are also shown.

During the information compilation activity, the right side of the workspace lists the task and its subdivision (pre-determined). One can move a snippet from the left side to the right side under an appropriate aspect by selecting that aspect from the drop-down box with that snippet and clicking on the right arrow. Since the workspace is updated in real time, such a change is reflected on both the participants' workspaces immediately.

4.2.3 Sessions

Each individual pair of participants came to the lab for two sessions that were one to two weeks apart. Each session lasted about one and a half hours. The flow for each session is depicted in Figure 4.7.

During the first session the participants were shown a video tutorial demonstrating the use of Coagmento and the process of collecting relevant information (snippets of



Figure 4.5: Coagmento workspace during information seeking part.

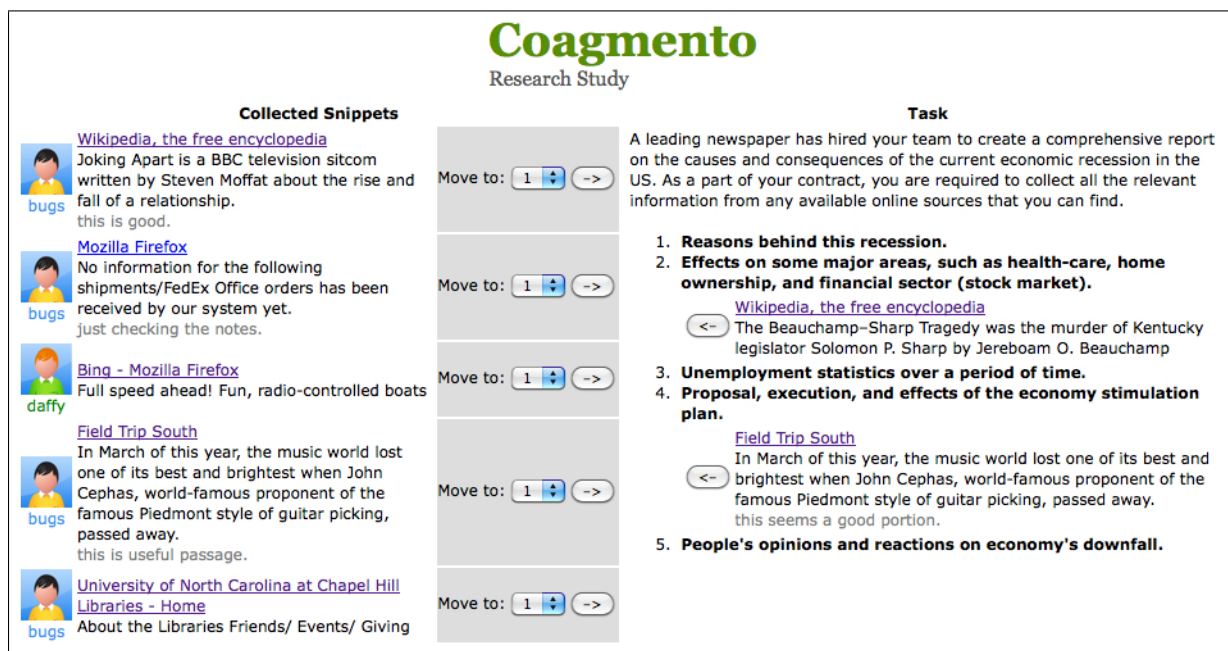


Figure 4.6: Coagmento workspace during information compilation part.

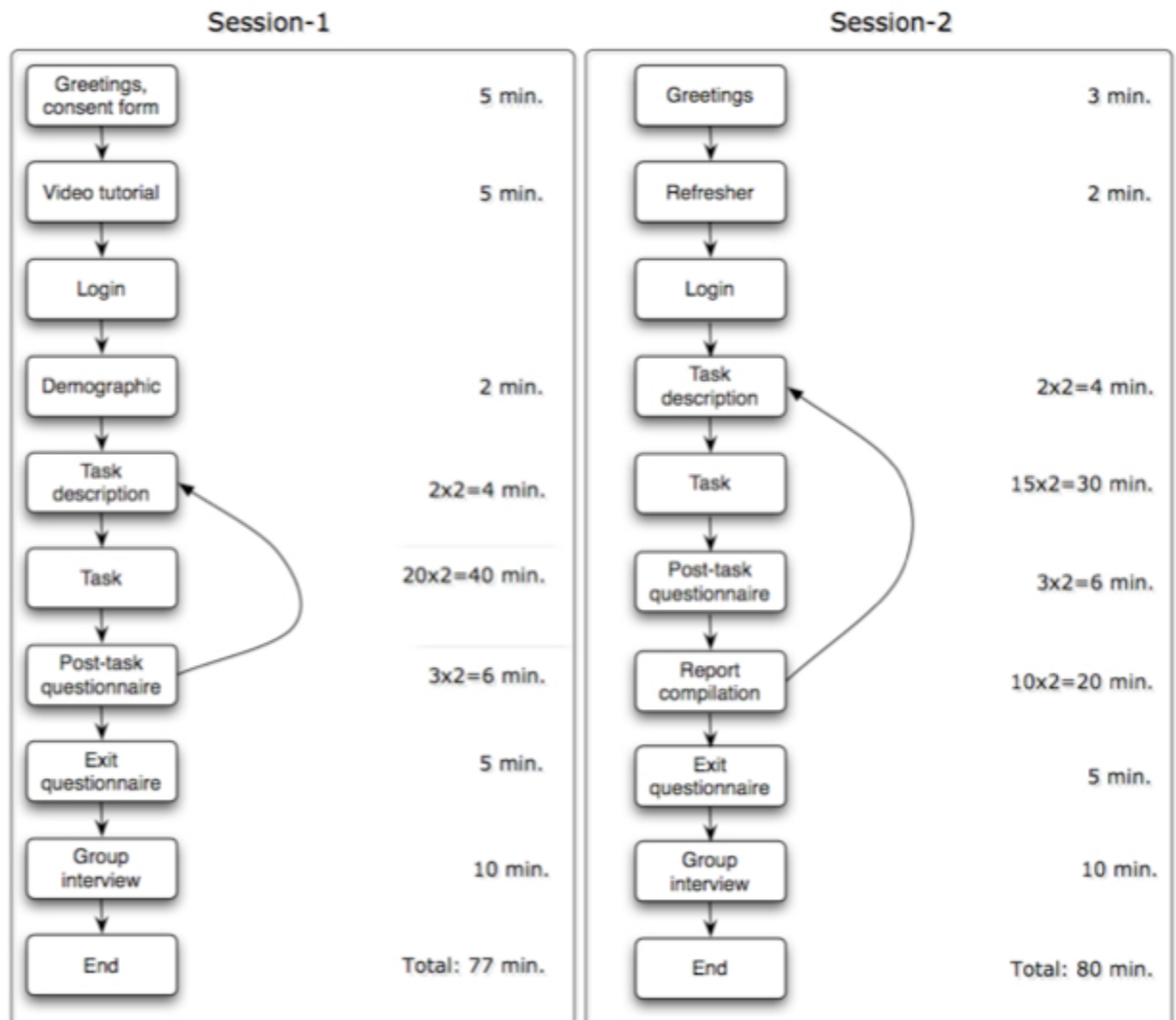


Figure 4.7: Flow of the sessions.

text). After the tutorial, the participants were placed in different rooms so that they could not talk to each other directly or see what the other person was doing (Figure 4.8). Both the participants used typical mid-end PC workstations, running Windows XP, with Ethernet connectivity and 19" monitors.

The supervisor (a researcher) took his place outside, stationed so that he could see both the participants. Once the participants logged in, they filled out a demographic questionnaire (Appendix D) and began working on the first task.



Figure 4.8: Participants during a study session.

As discussed below, the tasks were simulated work tasks and about 20 minutes into their work the supervisor sent out a message via the sidebar chat asking them to stop the task and fill in a set of online questionnaires asking about their progress on the task, as if their supervisor had requested an update. Once both the participants finished their individual questionnaires, they were asked to start working on the second task. The participants were once again interrupted about 20 minutes later, and asked

to complete the same questionnaires.

After the post-task questionnaire for the second task, the participants were asked to fill in an end of session questionnaire that included questions about ease of use and satisfaction, and engagement. After the participants submitted their individual end of session questionnaires, they were brought together and asked a few questions relating to their experience working in collaborating with Coagmento. These questions were based on the session and the condition they were in.

For the second session, the participants were given a refresher of the system and shown how to compile their final report by grouping their collected snippets into different categories for a given task. The categories were presented in the task statement and corresponded to different aspects of the work task (see task statements below). The participants were then asked to take their places in the room other than the one they had used the last time, to take care of any bias the participants may have for the machine or the room they used.

After 20 minutes of additional work on task one, they completed the post-task questionnaire with questions on task progress, awareness, and cognitive load, and were asked to organize their collected snippets by placing each relevant snippet in one of the categories for a given task. When they had finished organizing their snippets, they worked through the second task, including collecting their information, post-task questionnaire, and organizing the snippets. Upon completing both the tasks the participants completed an exit questionnaire, after which they were brought together and interviewed.

4.2.4 Tasks

The participants were asked to collect relevant information for two exploratory tasks that were designed to be realistic work tasks that might be of interest to the participant

pool (Borlund and Ingwersen, 1999). Rather than asking participants to create their own organizations for the pertinent snippets, the task statements identified specific issues that should be addressed and these issues were used as organizing bins for the collected snippets.

The task descriptions as given to the participants are provided below.

Task-1: Economic recession

“A leading newspaper has hired your team to create a comprehensive report on the causes and consequences of the current economic recession in the US. As a part of your contract, you are required to collect all the relevant information from any available online sources that you can find.

To prepare this report, search and visit any website that you want and look for specific aspects as given in the guideline below. As you find useful information, highlight and save relevant snippets. Later, you can use these snippets to compile your report. You may also want to save the relevant websites as bookmarks, but remember - your main objective here is to collect as many relevant snippets as possible.

Your report on this topic should address the following issues: reasons behind this recession, effects on some major areas, such as health-care, home ownership, and financial sector (stock market), unemployment statistics over a period of time, proposal, execution, and effects of the economy stimulation plan, and people’s opinions and reactions on economy’s downfall.”

Task-2: Social networking

“The College Network News Channel wants to do a documentary on the effects of social networking services and software. Your team is responsible for collecting various relevant information (including statistics) from the Web. As a part of your assignment,

you are required to collect all the relevant information from any available online sources that you can find.

To prepare this report, search and visit any website that you want and look for specific aspects as given in the guideline below. As you find useful information, highlight and save relevant snippets. Later, you can use these snippets to compile your report. You may also want to save the relevant websites as bookmarks, but remember - your main objective here is to collect as many relevant snippets as possible.

Your report on this topic should address the following issues: emergence and spread of social networking sites, such as MySpace, Facebook, Twitter, and del.icio.us, statistics about popularity of such sites (How many users? How much time they spend? How much content?), impacts on students and professionals, commerce around these sites (How do they make money? How do users use them to make money?), and examples of usage of such services in various domains, such as health-care and politics.”

4.2.5 Evaluation

While the focus of this study is understanding the impact of awareness in CIS, several constructs need to be measured to get the whole picture. These constructs are listed below along with their definitions and instruments used to measure them as appropriate. It is important to note that the majority of these constructs were evaluated considering an individual as the unit since all the questionnaires were submitted by every participant individually. A given questionnaire may also be presented to the same participant several times during their two tasks spanning two sessions, and all of their responses on multiple occurrences of the same questionnaire were counted (repeated measures), instead of averaging them, for analysis. Whenever appropriate, a group was considered as the unit. Construct-wise method of analysis is presented later.

- *Effectiveness*

Given that the tasks for this study require exploratory searching and there may not exist a clear set of relevant documents against which one can evaluate effectiveness of one's collected information, a proxy measure was created. First, assume that everything that all the teams found and saved for a given task is relevant, and then comparing each team's "recall" with respect to that common pool. Next, it is assumed that any document that one team found useful can be confirmed if another team also found it useful. Thus, the "precision" of a team was measured by computing how much of its collected information was found useful by at least one other team. These are not claimed to be the best criteria for measuring effectiveness, but they seem to be appropriate for the given situation.

$$\text{Recall of a team} = \frac{\text{No. of snippets collected by that team}}{\text{No. of snippets collected by all the teams}} \quad (4.1)$$

$$\text{Precision of a team} = \frac{\text{No. of overlapping snippets with at least one more team}}{\text{Total no. of snippets collected by that team}} \quad (4.2)$$

- *Efficiency*

The logs were examined to understand the amount of work each individual participant did during the study. Since each participant was given about the same amount of time for each task, this can serve as a way to measure efficiency.

Another way to measure efficiency for a team, is to look at the overlaps in the webpages that the participants of a given team had. Since the tasks are recall-oriented and time was limited, an efficient team will try to avoid duplicating each other's work and have little overlap in the webpages they view as possible.

- *Awareness*

This is at the core of this study and was measured in several ways. First, the participants were asked to report various data about their project when interrupted during the task using the following questionnaire.

Q1. How many webpages do you think you viewed?

Q2. How many webpages do you think your team viewed?

Q3. How many webpages do you think you saved?

Q4. How many webpages do you think your team saved?

Q5. How many snippets do you think you saved?

Q6. How many snippets do you think your team saved?

Q7. How many queries do you think you used?

Q8. How many queries do you think your team used?

Since each team-member reported these numbers about the project status independently, their reported numbers were compared to see how well they were coordinated. Of course, the team-members could communicate with each other at that time to stay on the same page, reporting their data. Thus, the amount of this communication that took place as well as the time spent filling in project status was also assessed. Their reported numbers were compared with actual statistics at that time as a way to find out how close their perceptions were to the reality.

In addition to measuring their perceived awareness about the project, direct questions related to various aspects of situational self-awareness, derived from (Govern

and Marsch, 2001), were asked on a questionnaire as shown below for the participants to rate on the scale of 1 to 7.

Q1. Right now, I am keenly aware of everything in my environment.

Q2. Right now, I am conscious of what is going on around me.

Q3. Right now, I am conscious of all objects around me.

Q4. Right now, I am concerned about what my teammate thinks of me.

Q5. Right now, I am aware of what my teammate just did.

Q6. Right now, I am conscious that my teammate is aware of my actions.

Q7. Right now, I am aware of how well we performed together in the team.

Finally, the participants were asked to rate (scale 1 to 7) two specific questions about their perceived personal and group awareness at the end of each session on an exit questionnaire.

Q1. I was aware of my own history (documents viewed and saved, snippets saved, queries used, etc.).

Q2. I was aware of my team's history (documents viewed and saved, snippets saved, queries used, etc.).

- *Effort*

Since each team received approximately the same amount of time, the physical effort can be measured by the number of actions taken in that time.

The mental effort or cognitive load was measured using a questionnaire derived from NASA's TLX instrument,¹ which was presented to each participant at the end of every task in both the sessions. The participants were asked to rate each of the following on the scale of 1 to 7.

Q1. How mentally demanding was this task? (Very low to Very high)

Q2. How physically demanding was this task? (Very low to Very high)

Q3. How hurried or rushed was the pace of the task? (Very low to Very high)

Q4. How successful were you in accomplishing what you were asked to do? (Perfect to Failure)

Q5. How hard did you have to work to accomplish your level of performance? (Very low to Very high)

Q6. How insecure, discouraged, irritated, stressed, and annoyed were you? (Very low to Very high)

- *Ease of use and satisfaction*

The participants were asked to rate (scale 1 to 7) several factors about the system at the end of each session as shown in the following questionnaire. The questionnaire was derived from the original Computer System Usability Questionnaire (Lewis, 1995),² removing those questions that were not relevant for this study. Responses to these questions shed light on their perceived ease of use and satisfaction.

¹Taken from http://www.cc.gatech.edu/classes/AY2005/cs7470_fall/papers/manual.pdf

²Also available from <http://oldwww.acm.org/perlman/question.cgi?form=CSUQ>.

-
- Q1. Overall, I am satisfied with how easy it is to use this system.
- Q2. I can effectively complete my work using this system.
- Q3. I am able to efficiently complete my work using this system.
- Q4. I feel comfortable using this system.
- Q5. It was easy to learn to use this system.
- Q6. I believe I became productive quickly using this system.
- Q7. It is easy to find the information I need.
- Q8. The information provided for the system is easy to understand.
- Q9. The organization of information on the system screens (toolbar, sidebar) is clear.
- Q10. The interface of this system is pleasant.
- Q11. I like using the interface of this system.
- Q12. This system has all the functions and capabilities I expect it to have.
- Q13. Overall, I am satisfied with this system.
-

- *Engagement*

This study is set in an interactive environment and one of the important factors to measure was the level of engagement the participants experienced while working in this environment. To measure this, the participants were presented with a questionnaire, taken from Ghani, Supnick, and Rooney (1991), at the end of each session to rate on the scale of 1 to 7.

Using the system was

- Q1. Uninteresting 1 2 3 4 5 6 7 Interesting
- Q2. Not Enjoyable 1 2 3 4 5 6 7 Enjoyable
- Q3. Dull 1 2 3 4 5 6 7 Exciting
- Q4. Not Fun 1 2 3 4 5 6 7 Fun

How did you feel while collaborating with this system

- Q5. Not absorbed intensely 1 2 3 4 5 6 7 Absorbed intensely
- Q6. Attention was not focused 1 2 3 4 5 6 7 Attention was focused
- Q7. Did not concentrate fully 1 2 3 4 5 6 7 Concentrated fully
- Q8. Not deeply engrossed 1 2 3 4 5 6 7 Deeply engrossed
-

These constructs, their definitions, and evaluation measures are given in Table 4.1.

The hypotheses relating to these constructs are presented in Table 4.2.

Table 4.1: Summary of constructs, their definitions, and evaluation measures. I = individual measure. G = group measure.

Construct	Definition	Evaluation
Effectiveness	Quality of the information (snippets) collected.	Recall and precision based on the pooled results. (G)
Efficiency	Amount of information discovered in given time.	Individual usage from logs. (I) Overlap of viewed webpages. (G)
Awareness	Perceived and reported consciousness of personal and group's history and progress.	Perceived - reality check for the project status. (I, G) Direct - Govern and Marsch's Situational Self-Awareness Scale (SSAS). (I)
Effort	Number of physical actions and perceived mental load.	Cognitive load (mental effort) - NASA's TLX. (I) Physical effort - number of physical actions taken. (I)
Ease of use and satisfaction	User's ratings of various system components.	Abbreviated CSUQ (I)
Engagement	User's involvement with the system as well as ongoing project.	Derived from Ghani, Supnick, and Rooney (1991) (I)

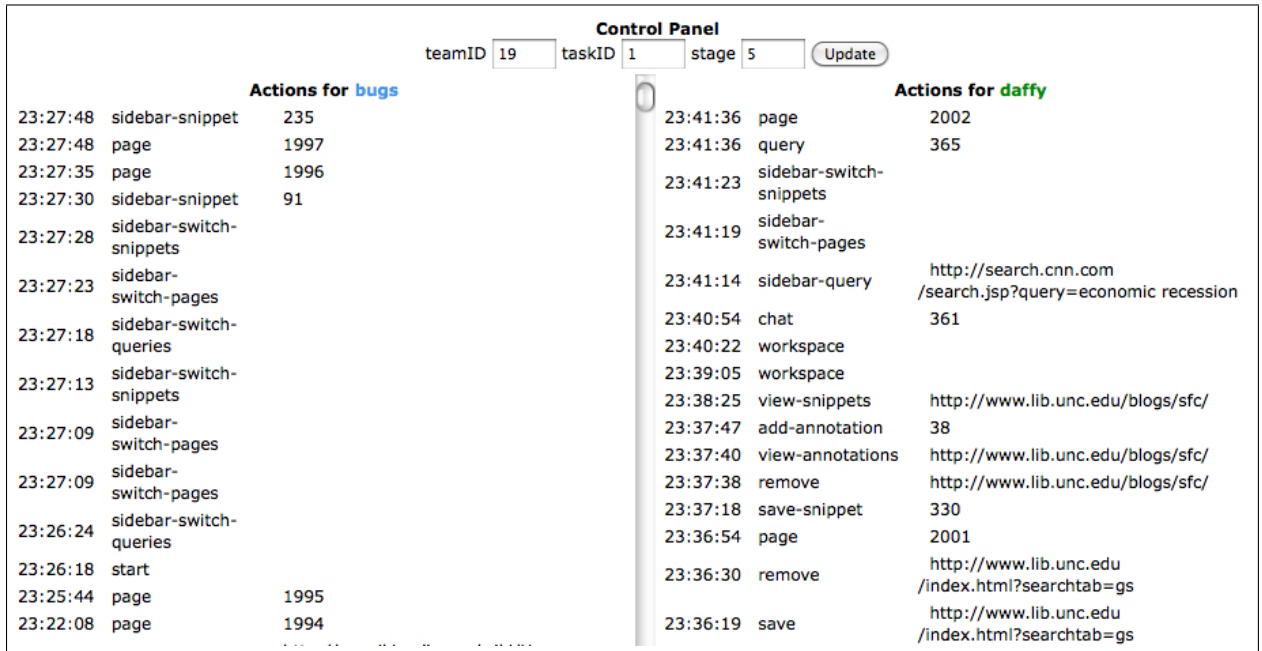
Table 4.2: Hypotheses related to various constructs and conditions.

Construct	Hypothesis
Effectiveness	$1 < 2 < 3$
Efficiency	$1 < 2 < 3$
Awareness	$1 < 2 < 3$
Effort - physical effort	$1 \sim 2 \sim 3$
- cognitive load	$1 \sim 2 \sim 3$
Ease of use and satisfaction	$1 \sim 2 \sim 3$
Engagement	$1 < 2 < 3$

4.2.6 Supervising the study

In order to supervise the study effectively, an interface was created that allowed me, as the supervisor, to monitor every action taken by either of the participants. This interface is displayed in Figure 4.9. As shown, individual actions of each of the participants are shown side-by-side and updated in real time. The actions include page view, saving a webpage, querying, collecting a snippet, making an annotation, using one of the objects from the sidebar (for conditions 2 and 3), and chatting. In addition to this, various actions related to questionnaires filling are also reported. This way the supervisor could know when a participant started filling in a questionnaire and when it was submitted. This was essential for the flow of the study since in order to move forward, the supervisor had to make sure the appropriate questionnaires were filled in by both the participants.

In addition to getting real-time updates of participants' actions, the supervising interface also allowed me to control the flow of a session. This was done through the control panel shown at the top of this interface (Figure 4.9). Using the control panel, the supervisor can indicate the team, task, and the stage at which the participants should be. The stages include: 1=demographic, 2=task, 3=post-task questionnaire, 4=exit questionnaire, and 5=report compilation.



During the study session, no verbal communication took place between the participants or the participants and myself. We used the sidebar chat for communicating with each other. As a supervisor, the only time I communicated was when I wanted to instruct the participants to do something, such as stopping the task and filling in a questionnaire (Figure 4.10).

4.3 Description of overall user activity

In this section, a detailed description of major objects that the participants worked with during the study is presented. These objects include visited and saved webpages, and search engine queries. The description is presented aggregating actions of all the participants irrespective of their condition, team, or session.³ The data is reported separately for individual tasks or sessions when appropriate. Further description of the

³For non-aggregated data analysis, see the next section.

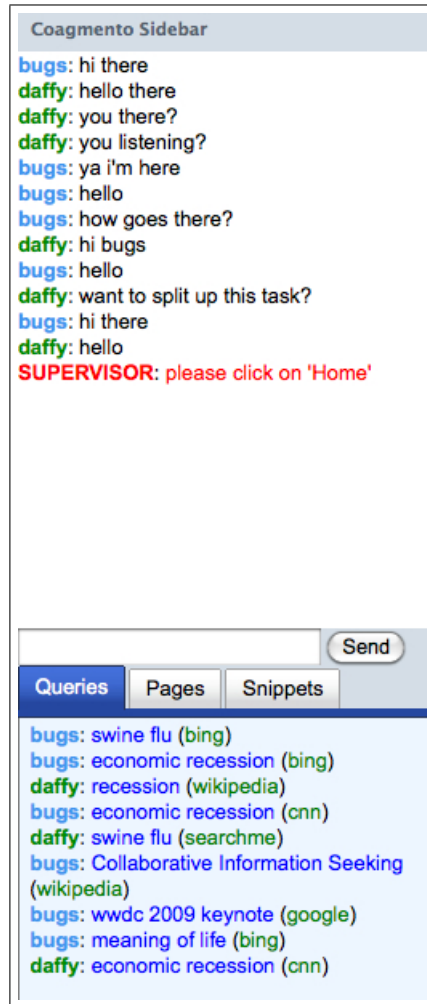


Figure 4.10: Example sidebar showing supervisor's message.

data, focusing on query re-usage, is given in Appendix E.

4.3.1 Webpages and sources

Table 4.3 and 4.4 list the top sites that the users visited for each task. Google was the most visited site for both the tasks, followed by news sites, such as the New York Times and CNN. Wikipedia was also one of the top sources for both the tasks. For Task-1, the participants also visited the Bureau of Labor Statistics (BLS) and the Recession.org website, where many up-to-date statistics on the current economic recession can be

obtained. Mining the log data revealed that these websites were discovered primarily due to queries such as “unemployment statistics US” and “economic recession” rather than by following links or directly typing in URLs.⁴

4.3.2 Search queries

The participants used a total of 4263 queries (aggregated over both the tasks and sessions), of which 1621 were source-wise unique queries (unique source-query pair), and 1536 were overall unique queries. Thus, nearly 40% of the queries issued were repeated at least once (by the same or some other participants). Table 4.5 reports the session-wise usage of search services. Note that, for the first session, the default search engine in the Firefox search box (available in the top right corner) was set to Google, and was changed to Bing for the second session. It is possible that, due to this switch, Bing’s usage was found to be much higher during the second session.

Tables 4.6 and 4.7 show the sources that were queried by the participants (total and unique queries). This information is aggregated in Figure 4.11. As we can see, a significant portion (more than 80%) of all the queries were sent to Google, with CNN (primarily for Task-1, about 5%) and Bing (about 10%) next most used. It is interesting to note that, with every source, a large portion of queries were repeats.

4.4 Analysis

The focus of this study is to understand how different kinds of awareness supported by the system could impact different elements and processes of a collaborative project. As discussed in the study design, the participants were elicited several times during the two

⁴A few times participants even typed queries such as “bls” in Google indicating participants’ prior familiarity with this source, but not remembering its actual address.

Table 4.3: Top 40 sources for Task-1, economic recession.

Source	Times visited
google	1809
cnn	293
bing	206
nytimes	200
bls	183
wikipedia	164
recession	152
wsj	93
economist	79
about	77
recovery	74
reuters	65
washingtonpost	63
yahoo	59
msn	56
unc	53
businessweek	52
ebscohost	48
forbes	48
bea	47
usnews	41
blogspot	40
usatoday	37
serialssolutions	35
npr	28
answers	25
lexisnexis	24
time	24
worldbank	23
morebusiness	22
ezinearticles	22
cepr	22
cbsnews	21
gallup	19
associatedcontent	18
heritage	16
cqpress	16
huffingtonpost	15
pollingreport	15
rgemonitor	14

Table 4.4: Top 40 sources for Task-2, social networking.

Source	Times visited
google	2235
bing	279
wikipedia	231
nytimes	143
facebook	87
cnn	85
yahoo	79
go	59
web-strategist	56
delicious	52
techcrunch	50
mashable	48
ask	45
time	44
socialnetworkingwatch	40
twitter	38
msn	38
newsweek	37
associatedcontent	36
businessweek	36
wordpress	31
ebscohost	31
marketingpilgrim	29
blogspot	27
ezinearticles	27
economist	26
forbes	24
usatoday	24
indiana	23
unc	23
toptenreviews	22
post-gazette	21
alexa	21
myspace	20
istrategylabs	20
about	20
readwriteweb	20
wsj	20
slideshare	19
ignitesocialmedia	19

Table 4.5: Search services used for each session.

Source	Session-1	Session-2
altavista	5 (0.24%)	3 (0.13%)
aol	2 (0.09%)	39 (1.73%)
ask	4 (0.19%)	40 (1.78%)
bing	17 (0.83%)	455 (20.20%)
cnn	100 (4.88%)	24 (1.07%)
google	1840 (89.84%)	1667 (74.02%)
msn	7 (0.34%)	5 (0.22%)
wikipedia	2 (0.09%)	3 (0.13%)
yahoo	71 (3.47%)	16 (0.71%)
Total	2048 (100%)	2252 (100%)

Table 4.6: Usage of search services used for Task-1.

Source	Total queries	Unique queries
google	1532	582
bing	199	99
cnn	94	42
yahoo	34	15
msn	7	3
altavista	3	2
ask	3	1
wikipedia	3	3
Overall	1875	702

Table 4.7: Usage of search services used for Task-2.

Source	Total queries	Unique queries
google	1977	750
bing	273	91
yahoo	53	15
ask	40	12
cnn	30	14
altavista	5	2
msn	5	1
aol	2	1
wikipedia	2	2
Overall	2388	851

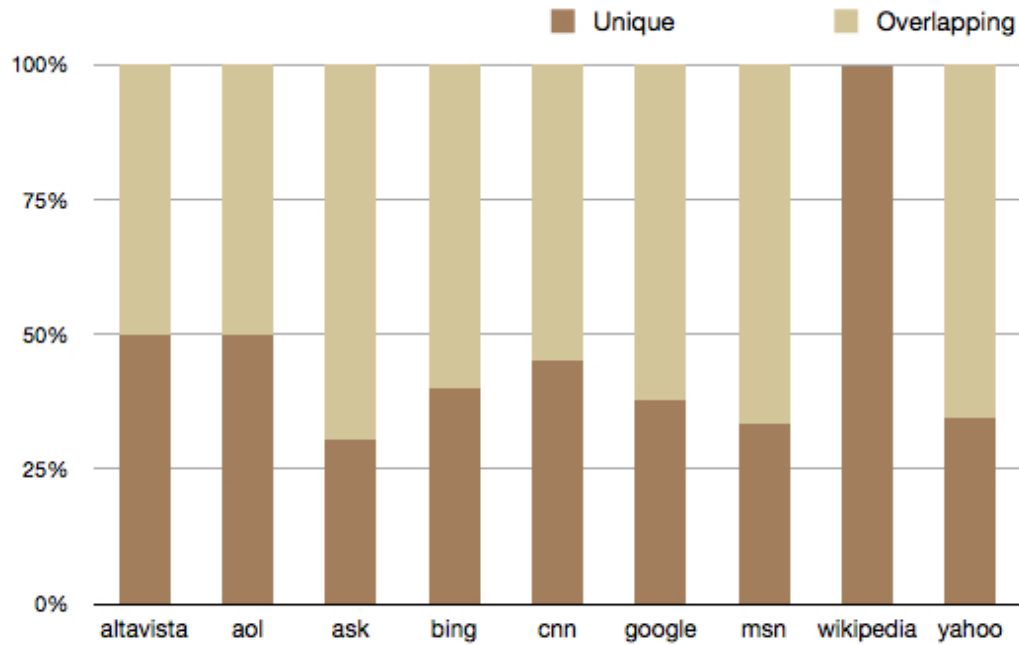


Figure 4.11: Source-wise query usage and overlap in aggregation.

sessions for input on various factors, including awareness, cognitive load, ease of use, user satisfaction, and engagement. This section presents an analysis of participants' awareness (perceived and real) along with these other factors.

Note that the primary independent variable is the kind of awareness support provided, which is reflected in the three conditions created. For most of the analyses, no significant difference was found between two tasks or two sessions among the three conditions (see Appendix F) . Therefore, in most cases, the analysis is shown with tasks and/or sessions collapsed, thus, measuring the difference among the three conditions only.

The method used for the analysis presented here is depicted in Figure 4.12. For a questionnaire, if creating an index that combines all of the questions on that questionnaire was appropriate, that index's statistical reliability is checked and if reliability statistic $\alpha > 0.7$, such an index is created by averaging the responses for all the questions. One-way ANOVA on this index is done, and if found significant with $p < 0.05$, Sheffe's post-hoc test was executed. If creating an index was not appropriate or its statistical reliability was not sufficient, or if an item for analysis was log data, straight one-way ANOVA was done, followed by Sheffe's post-hoc test, if appropriate.

4.4.1 Effectiveness

Each team's effectiveness in this study was measured using recall and precision (see Equations 1 and 2) of the snippets they collected. As shown in Table 4.8, no difference was found among the conditions for either of the measures.

Unusually low recall values can be attributed to the way recall was computed here. The 42 teams that participated collected total 3278 snippets, averaging about 78 snippets per team. Thus, on average, a team's recall would be $78/3278=0.023$. Further investigations into measuring the quality of collected information, that goes beyond

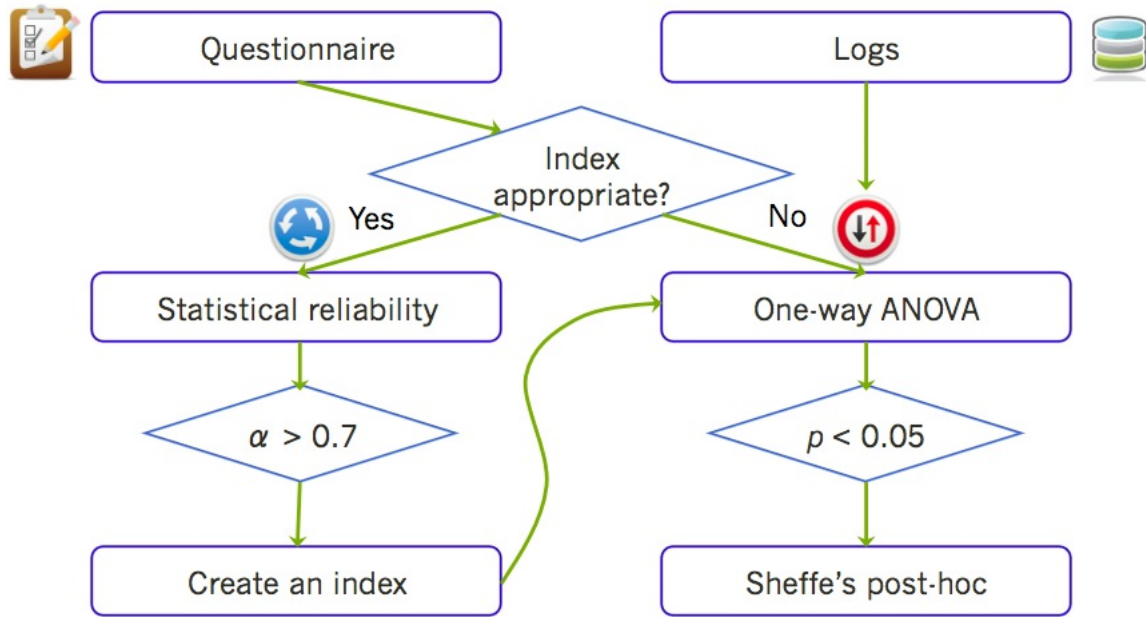


Figure 4.12: Method for statistical analysis of the data.

simple recall and precision, is needed.

Table 4.8: Quality of collected snippets as measured by recall and precision. For one-way ANOVA, between groups $df = 2$ and within groups $df = 39$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Recall ($F = 0.086, p = 0.918$)	0.024	0.014	0.023	0.012	0.024	0.007
Precision ($F = 1.111, p = 0.339$)	0.474	0.147	0.409	0.129	0.399	0.155

4.4.2 Efficiency

Efficiency of each team was measured in two different ways: looking at each participant's activities (Table 4.9), and finding the webpages overlapped between the collaborators in a given pair (Table 4.10).

Participants in condition 3 used significantly more unique queries than those in

Table 4.9: Summary statistics of different activities per person. For one-way ANOVA, between groups $df = 2$ and within groups $df = 81$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Total queries ($F = 2.391, p < 0.098$)	44.50	20.37	49.61	18.46	58.07	29.92
Unique queries* ($F = 3.177, p < 0.0487$)	24.11	10.37	29.11	9.94	31.93	14.44
Total webpages viewed ($F = 1.952, p < 0.149$)	124.71	31.22	142.86	45.10	144.64	47.15
Unique webpages viewed ($F = 2.788, p < 0.067$)	79.43	16.95	92.89	27.00	92.61	27.68
Webpages saved ($F = 0.489, p < 0.615$)	13.54	5.75	14.86	8.04	15.54	8.95
Snippets collected ($F = 0.141, p < 0.868$)	19.95	14.73	18.79	13.07	19.80	9.38

*Difference across conditions was statistically significant ($p < 0.05$).

Table 4.10: Summary statistics of webpages overlapped between the collaborators in a given pair. For one-way ANOVA, between groups $df = 2$ and within groups $df = 39$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Webpages overlapped within team ($F = 0.036, p < 0.964$)	4.00	3.96	4.29	2.27	4.07	2.16

condition 1 ($p < 0.05$). Given that there was no difference between these two conditions for the total number of queries used, we can say that the participants in condition 3 had much less overlap in their query usage, and could try more unique searches for the task at hand.

Looking at the overlap of webpage visits between the two participants in a team (Table 4.10), we see that those in condition 2 had slightly more overlap than the other two conditions. However, these differences were not found to be statistically significant.

These observations inform us that even though it appears that the teams in each condition are doing the same amount of work, condition 3 teams are clearly able to explore more venues of information overall. This can be attributed to the kind of awareness provided. For example, a participant in condition 1 has no support for knowing the search queries he or his teammate used in the past. Therefore, he is likely to repeat a search even when it is not needed. A participant in condition 2, on the other hand, can avoid repeating his own searches unless necessary due to the personal peripheral awareness support. A participant in condition 3 will have knowledge about his as well as his teammate's searches. This explains why greater differences were found between conditions 1 and 3 for unique query usage, and not among other pairs of conditions.

Another important aspect of the assigned tasks was the snippets. As shown in Table 4.9, no significant difference was found in the number of snippets collected by the teams in different conditions. Much of the snippets collecting can be attributed to personal motivation. The team that collected the most number of snippets was in condition 1. Given that time was limited, the tasks exploratory, and a prize to be won for those collecting many snippets, this team quickly divided the task at hand and went through as many webpages as possible collecting numerous snippets. This team gathered a total of 223 snippets combined for both the tasks during the two sessions.

This was significantly higher than the average, which was about 78. The minimum number of snippets collected was 38 by a team in condition 1.

To summarize, it is hard to claim difference in efficiency among different conditions. On the other hand, through query usage, it was found that teams in condition 3 were being more efficient in their work, that is, they managed to explore more volume of information with the same amount of time and work.

4.4.3 Awareness

Since awareness was the core issue of this investigation, much attention was given to measuring it in a number of ways. To measure how much the participants were aware of various factors during the experiment, they were presented with several questionnaires at different points.

Perceived awareness

In order to obtain feedback on how aware the participants felt, various statements, inspired by Govern and Marsch (2001), were presented to the participants at the end of each task for both the sessions. The participants were asked to select a value between 1 (Strongly disagree) and 7 (Strongly agree) for these statements.

Since creating an index is appropriate for this questionnaire, and it was found to be statistically reliable with $\alpha = 0.781$, an index was created by averaging the responses on all the seven questions for every individual. Table 4.11 reports various statistics for this index across all the conditions. No significant difference was found among the three conditions. This indicates that as far as being conscious of the situation around them was concerned, participants in each condition were similarly aware.

In addition to these post-task questions, the participants were also asked to rate two specific questions relating to personal and group awareness at the end of each

Table 4.11: Index for perceived situational awareness questions. For one-way ANOVA on each question, between groups $df = 2$ and within groups $df = 333$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Index for perceived situational awareness ($F = 1.653, p < 0.193$)	4.69	0.859	4.56	1.062	4.79	0.924

session. The average responses on the scale 1 (strongly disagree) to 7 (strongly agree) are shown in Table 4.12 for each condition. Condition 2 participants reported higher personal awareness (Q1), even higher than those in condition 3 ($p < 0.018$). This is not surprising given that condition 2 had exclusive support for personal history. But when it comes to being aware of the team, condition 3 participants reported significantly higher values than those in condition 1 ($p < 0.001$) or 2 ($p < 0.001$). This is, once again, an indication that the participants in condition 3 were much more aware of the status of the team and the project throughout their collaborative endeavor.

Table 4.12: Summary statistics for responses on perceived awareness questions at the end of each session. For one-way ANOVA on each question, between groups $df = 2$ and within groups $df = 165$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Q1. personal awareness* ($F = 4.426, p < 0.013$)	5.73	0.842	6.12	0.955	5.59	1.141
Q2. group awareness* ($F = 19.221, p < 0.001$)	4.55	1.205	3.95	1.531	5.50	1.250

*Difference across conditions was statistically significant ($p < 0.05$).

The questions listed above asked the participants to report their level of awareness about certain aspects directly. This may not be enough to tell the whole story about awareness. The participants, therefore, were asked more specific questions about the status of their ongoing project.

Project status awareness

As described earlier, the participants were interrupted after about 20 minutes for a given task and asked to report the project status. This status was obtained by asking specific questions about viewed and saved webpages, used queries, and collected snippets, both individually and as a team.

The first thing to analyze from the responses of these questions is whether the collaborators of a given team were in agreement while reporting the status of the project. These questions are Q2, Q4, Q6, and Q8 in the questionnaire given earlier in the method section. It was found that creating an index was appropriate and statistically reliable ($\alpha = 0.759$). Table 4.13 reports the statistics with this index for average differences (absolute values) in the responses to these questions for a given team. As we can see, the teams in condition 3 have the least difference for these questions, and was significant compared to condition 1 ($p < 0.024$) and condition 2 ($p < 0.001$). In other words, these results indicate that those in condition 3 were highly coordinated when it came to being aware of the group's history and progress. It is also interesting to note is that even teams in condition 1 had smaller differences than those in condition 2 ($p < 0.046$).

Table 4.13: Index for absolute difference between two collaborators' reported numbers for group's status. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Index for group reporting difference* ($F = 13.827, p < 0.001$)	2.38	3.247	3.99	4.896	0.598	0.756

*Difference across conditions was statistically significant ($p < 0.05$).

Since the participants were allowed to chat with each other while filling in these values, we cannot simply look at their differences and comment about the strength of their collaboration. We also need to consider the cost associated with achieving the

coordination. This was done by analyzing the time and the number of messages that they exchanged while completing this questionnaire.

The results are reported in Table 4.14. It turns out that there was no statistically reliable difference between conditions 1 and 2 for the amount of time used, but those in condition 1 exchanged significantly more messages than those in condition 2 ($p < 0.018$). It was also found that condition 3 had significantly less overhead in terms of communication compared to conditions 1 ($p < 0.001$) and 2 ($p < 0.001$) while filling in this status report. They also spent less time compared to condition 1 ($p < 0.001$) and 2 ($p < 0.001$). This shows that the participants in condition 3 were well aware of the status of their projects and could report it without spending much time or communication.

Table 4.14: Average amount of time spent and messages exchanged by each participant while reporting project status. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Avg. time (sec)* ($F = 1.714, p < 0.001$)	198.18	114.149	180.12	130.626	123.04	71.000
Avg. no. of messages* ($F = 39.981, p < 0.001$)	13.94	11.963	10.12	10.576	2.37	6.103

*Difference across conditions was statistically significant ($p < 0.05$).

Perceived awareness vs. reality

Next, let us look at how close participants' reporting about project awareness was to the reality. For this purpose, eight questions on which a participant was asked to report various numbers regarding personal or team history on the task at hand were considered. The actual numbers were then calculated using the log data and compared to the reported numbers. The results of this comparison are presented in Tables 4.15 and 4.16. The following observations can be made from these results.

- The participants in condition 2 did significantly better than condition 1 while reporting personal numbers on webpages viewed ($p < 0.001$), webpages saved ($p < 0.001$), snippets collected ($p < 0.001$), and queries used ($p < 0.018$). In fact, condition 2 also did better than condition 3 on webpages viewed ($p < 0.001$), snippets collected ($p < 0.005$), and queries used ($p < 0.001$). This makes sense, as they were the only ones with exclusive support for personal awareness.
- However, the participants in condition 2 did not have group awareness support, nor did those in condition 1. Therefore, the participants in condition 3 could report the numbers for group progress that was significantly closer to the reality compared to those reported by condition 1 or 2. Condition 3 outperformed condition 2 on webpages viewed ($p < 0.017$) and saved ($p < 0.001$), snippets collected ($p < 0.008$), and queries used ($p < 0.030$). They also did better than those in condition 1 on all of these fronts ($p < 0.001$).
- The participants in condition 3 did significantly better even on personal awareness related questions on webpages viewed ($p < 0.001$) and saved ($p < 0.013$) compared to those in condition 1, the participants in condition 3 did not have explicit support for personal awareness, they were in a better position to make educated guesses (and sometimes actual counts) using the group awareness information, than those in condition 1.

It is clear from these observations that the condition 2 interface provides the most suitable support for personal awareness, whereas condition 3 has the best support for group awareness. An ideal interface may include group information with the ability to switch to personal information as needed. In other words, such an interface would provide more control for the awareness to the user.

Table 4.15: Average difference between reported numbers and real numbers for questions on personal progress. For one-way ANOVA on each response, $df = 2$ and within groups $df = 333$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Webpages viewed* ($F = 161.194, p < 0.001$)	16.68	10.855	0.35	1.486	5.65	4.960
Webpages saved* ($F = 11.331, p < 0.001$)	2.44	2.660	1.12	1.693	1.61	1.812
Snippets saved* ($F = 11.832, p < 0.001$)	2.86	4.721	0.62	2.023	2.16	3.349
Queries used* ($F = 23.192, p < 0.001$)	2.29	2.399	0.46	1.314	1.83	2.380

*Difference across conditions was statistically significant ($p < 0.05$).

Table 4.16: Average difference between reported numbers and real numbers for questions on group progress. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Webpages viewed* ($F = 198.123, p < 0.001$)	31.75	18.319	6.25	9.728	1.56	4.152
Webpages saved* ($F = 49.665, p < 0.001$)	4.27	3.435	3.66	3.335	0.72	1.209
Snippets saved* ($F = 29.554, p < 0.001$)	5.95	7.789	2.72	4.752	0.49	1.599
Queries used* ($F = 30.042, p < 0.001$)	4.32	4.215	2.09	3.639	0.91	1.585

*Difference across conditions was statistically significant ($p < 0.05$).

Workspace awareness

In each of the conditions, the participants were given the same kind of workspace (Figure 4.5). This workspace included the pages and the snippets saved by the team (both the participants). Thus, even the participants in condition-1, who had no personal or group awareness, could access this workspace and see where the team stood at a given point. Given that the workspace was a common component among all the conditions, it would be interesting to see if having peripheral (personal or group) awareness would change how often one accesses this workspace.

It was found that on average (Table 4.17), the participants in condition 1 accessed the workspace more than those in condition 2 ($p < 0.023$), and in condition 3 ($p < 0.001$). There was no statistically significant difference for workspace accesses between conditions 2 and 3. This indicates that the participants without any kind of awareness support had to access the common space more often to stay updated about personal and team progress, thus, further providing support for the usefulness of the peripheral awareness in collaborative projects.

Table 4.17: Average number of times workspace accessed by each participant. For one-way ANOVA on the index, $df = 2$ and within groups $df = 81$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Avg. no. of times workspace accessed* ($F = 12.818, p < 0.001$)	11.07	4.883	7.93	4.027	5.43	3.511

*Difference across conditions was statistically significant ($p < 0.05$).

4.4.4 Effort

In Section 4.2.5, two ways were proposed to measure the effort exerted by participants during this study: by analyzing physical actions taken, and by asking the participants to report the cognitive load they felt. Both of these analyses are presented below.

Physical effort

In order to measure the amount of physical work one did during this study, the logs were mined and the key actions that every team performed were enumerated. These key actions are: viewing or saving a webpage, executing a query, collecting a snippet, making or viewing an annotation, clicking on any item in the sidebar, and accessing the workspace. A summary of the number of such actions for each condition is reported in the second column of Table 4.18. As we can see, those in condition 3 performed more actions than those in condition 1 ($p < 0.046$).

Table 4.18: Average number of key actions, chat actions, and combined actions taken by each team. For one-way ANOVA on the index, $df = 2$ and within groups $df = 39$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Key actions* ($F = 3.393, p = 0.044$)	460.79	78.861	529.50	95.191	571.14	152.191
Chat actions* ($F = 7.592, p = 0.002$)	219.50	81.408	166.36	79.369	103.71	75.209
Combined actions ($F = 0.097, p = 0.908$)	680.29	115.133	695.86	92.494	674.86	172.400

*Difference across conditions was statistically significant ($p < 0.05$).

One important action that is omitted from these key actions is chat. Let us first look at chat action itself. The third row in Table 4.18 shows the average number of

chat messages sent between the teammates for each condition. As we see, now the relationships between different conditions are reversed. The participants in condition 1 exchanged significantly more messages than those in condition 3 ($p < 0.002$). It was already shown (Table 4.14) that the participants in conditions 1 and 2 exchanged significantly more chat messages while reporting the team's progress. The observations reported in Table 4.18 complements it by indicating a significantly higher volume of communication for these two conditions throughout the sessions. The implications of exchanging higher number of messages are difficult to state, but later in this chapter we will see if the greater amount of communication resulted in higher level of satisfaction or engagement among the participants.

Now, if we include chat into the other key actions (row four), we find that different conditions are not very far from each other in terms of the number of actions performed.

Synthesizing the observations from these three columns, we can say that while the teams in different conditions took similar number of key actions, those in condition 3 explored more information (and probably reflect more on that information as well), and spent less effort in communicating with each other. Of course, not all communication can be considered as cost, but as we saw before in Section 4.4.3, the teams in condition 1 and 2 sent significantly more messages (and spent more time) than those in condition 3 trying to coordinate with and update each other.

Cognitive load

One of the questionnaires that was presented to the participant after every task in both the sessions, was based on NASA's TLX for measuring the perceived cognitive load.⁵

The participants were asked to rate each of the six questions where '1' was very low and '7' was very high. An index was created for these questions ($\alpha = 0.730$), and is

⁵The actual instrument was taken from http://www.cc.gatech.edu/classes/AY2005/cs7470_fall/papers/manual.pdf

reported in Table 4.19.

Table 4.19: Index for cognitive load questionnaire filled by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 333$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Index for cognitive load questions ($F = 2.178, p = 0.115$)	3.21	0.825	3.19	0.802	3.41	1.005

No statistically significant differences were found among the three conditions. This informs us that, even those with a more complex interface such as condition 3, felt no more mentally loaded than those with a simple interface such as condition 1. This is good news for system designers as it indicates that we could provide additional support for CIS without adding any burden to the user.

4.4.5 Ease of use and satisfaction

At the end of each session, the participants were asked to rate a list of general questions about the interface common across all the three conditions, each on the scale of 1 (strongly disagree) to 7 (strongly agree).

An index was created for the 13 questions ($\alpha = 0.986$), which is reported in Table 4.20. It is interesting to see that compared to condition 2, participants in condition 1 reported higher values ($p < 0.010$). One of the implications of these results is that, while the participants in both the conditions liked the interface provided to them, they had differences for the expected functionalities and satisfaction they had from the system. Since they had personal awareness support, they realized how valuable it could be to have similar awareness about their teammates.

No difference was found between conditions 1 and 3, and 2 and 3. This, once again, indicates that a complex interface such as condition 3 can do equally well as a simple

Table 4.20: Index for ease of use and satisfaction questions responded by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Index for ease of use and satisfaction* ($F = 5.065, p = 0.007$)	5.99	0.661	5.04	1.789	5.31	2.120

*Difference across conditions was statistically significant ($p < 0.05$).

interface such as condition 1, when it comes to providing ease of use and satisfaction. Even though condition 3 had more information on the interface, the organization of that information was similar to condition 2 (on the periphery). This allowed the participants in condition 3 to carry on their tasks without being distracted, and at the same time feeling aware of the status of their teammates and the overall project. In other words, condition 3 interface could provide a better support for awareness without sacrificing ease of use and satisfaction.

4.4.6 Engagement

Since this study was done in an interactive environment, one important factor to analyze was how engaged the participants felt while working on their projects. To obtain this feedback, the participants were presented with a questionnaire at the end of each session, which was adopted from the full instrument presented in (Ghani, Supnick, and Rooney, 2007).

An index was created for these eight questions ($\alpha = 0.889$), and reported in Table 4.21. It was found that the participants in condition 3 gave significantly higher ratings compared to those in condition 2 ($p < 0.015$). This indicates that, even though the condition 3 interface was more complex, the participants felt more engaged while working in collaboration with the system.

Table 4.21: Index for engagement questions responded by individuals. For one-way ANOVA on the index, $df = 2$ and within groups $df = 165$.

	Condition 1		Condition 2		Condition 3	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
Index for engagement* ($F = 4.344, p = 0.014$)	5.58	0.719	5.35	0.886	5.81	0.825

*Difference across conditions was statistically significant ($p < 0.05$).

These results indicate that engagement in a user-centric CIS environment depends on not only the method of providing awareness (here, peripheral awareness), but also the nature of that awareness (personal vs. group). Condition 3 mirrored condition 2 in terms of the design for providing awareness (at the periphery), but the kind of awareness provided in the same screen space was different (group, instead of personal). This allowed the participants in condition 3 to carry on their tasks without being distracted, and at the same time feeling aware of the status of their teammates and the overall project.

4.5 Conclusion

This chapter presented the core investigation of this dissertation. It was focused on obtaining feedback on an experimental CIS system called Coagmento, and understanding various forms of user behavior while working in a CIS environment. But more importantly, this chapter presented a description and analysis for a number of core issues and challenges that arise in the domain of user-driven, computer-mediated, active, intentional, interactive, and mutually beneficial collaborative information seeking endeavors that span multiple sessions.

There are several lessons learned from the experiments and results reported in this chapter. They can be enumerated as the following.

- From the data analysis, it was discovered that, while there were no significant differences between the three conditions in terms of total amount of work the participants did, those in condition 3 were able to explore more unique information while keeping the communication needed to coordinate to a minimum. This observation supports the hypothesis that providing group information on the interface as peripheral awareness helps one to be efficient in exploring information.
- The participants in condition 3 not only reported higher awareness than those in other conditions, but their status reports were significantly closer to realities reflected by logged behavior. In addition, it was also found that the participants in conditions 1 and 2 exchanged significantly more messages and took significantly longer while providing their status reports. This shows that, in the absence of proper awareness support, the group members have to waste their resources to be on the same page, and they may be far off from what is the real status of their project.
- Having ready access to group information on the periphery helps one to be aware of the status of one's collaborators, as well as the task at hand, without being distracted.
- The three interfaces designed for the three conditions for this inter-participants study were in increasing order of complexity from 1 to 3. However, the participants found them relatively easy to learn. Those in condition 1 were found to be more satisfied than those in condition 2.
- Even though the condition 3 interface was much more complex than the condition 1 interface, the users were quite happy with it, and reported high levels of satisfaction, close to that of the condition 1 with their very simplistic interface. This demonstrates that we could provide the support for group information, which is

highly desired, with similar level of ease of use and satisfaction as the simplest possible system.

- Having real-time access to their teammate's actions and history made the participants in condition 3 more engaged than those in condition 1 or 2.
- A major usage of searching was for navigational purposes. For instance, often a participant would enter queries such as 'bls' or 'delicious' in Google to go to their homepages.

A summary of all the findings, corresponding to the hypotheses presented in Table 4.2 is provided in a very simplistic condensed form in Table 4.22. For each finding, the best performing condition is highlighted with a bold number. The last column of this table lists the tables and the corresponding page numbers on which the details of those findings can be found.

The findings reported in this table are based on the statistically significant relationships found among the conditions. For instance, regarding effectiveness and efficiency, no significant differences were found among the three conditions, and thus the findings for those constructs are reported as 1~2~3. Condition 2 was found to be better for personal awareness, and thus, it is given in bold case. Similarly, condition 3 outperformed both the other conditions for group awareness. Note that the issue of awareness is not that simple to summarize the findings that easily in this table, and other analyses presented earlier in this chapter should be considered while comparing the three conditions.

In general, the findings from this study strongly demonstrate that basic group awareness components offer substantial advantages to people working in CIS without adding new burdens to users.

Table 4.22: Summary of construct, hypotheses, and related findings. The best performing condition, if any, is given in bold letter.

Construct	Hypothesis	Finding	Tables
Effectiveness	$1 < 2 < 3$	$1 \sim 2 \sim 3$	4.8 (p.183)
Efficiency	$1 < 2 < 3$	$1 \sim 2 \sim 3$	4.9, 4.10 (p.184)
Awareness	$1 < 2 < 3$	Personal: $(1 \sim 3) < \mathbf{2}$ Group: $(1 \sim 2) < \mathbf{3}$	4.12 (p.187)
Effort - physical effort	$1 \sim 2 \sim 3$	$1 \sim 2 \sim 3$	4.18 (p.193)
- cognitive load	$1 \sim 2 \sim 3$	$1 \sim 2 \sim 3$	4.19 (p.195)
Ease of use and satisfaction	$1 \sim 2 \sim 3$	$\mathbf{1} > 2, 2 \sim 3, 1 \sim 3$	4.20 (p.196)
Engagement	$1 < 2 < 3$	$1 \sim 2, 2 < \mathbf{3}, 1 \sim 3$	4.21 (p.197)

Chapter 5

Discussion

Collaboration is a useful and often necessary component of several complex projects we work with every day. These projects range from planning a family vacation to co-authoring an article, and from filing taxes to cooking holiday meals. It is in human nature to collaborate with others when the task at hand is difficult or impossible to carry out by one. We have seen this nature from the days of hunting (Lee and DeVore, 1968) to the modern office environment (Hansen and Jarvelin, 2005). Given this, it is no surprise that in today's world, where information access, organization, and flow are considered to be essential to an individual or an organization's success, we find ourselves with a need to collaborate with others. This dissertation addressed some of the core issues of collaborative information seeking (CIS) in a computer-mediated environment.

This work started with the acknowledgement that, while collaboration is often desired or required, it may not always be necessary or cost-effective. With this, it was proposed to study the kind of collaboration that is explicitly defined or intentional, interactive, and mutually beneficial. Such collaboration is also envisioned being carried out actively by the participants over a number of sessions. With this notion of collaboration, a research agenda was set that included the following specific aspects of

CIS:

1. Understand instances, motivations, and methods for collaboration.
2. Develop systems and services to support CIS.
3. Investigate the impact of some of the core components of such systems on CIS projects, and user behavior in such environments.

A summary of the lessons learned for these aspects, in order to address the research questions provided in Chapter 1, is provided below, concluding with a description of a framework for supporting user-centric CIS.

5.1 Instances, motivations, and methods for collaboration

From several personal interviews it was discovered that there are many situations that naturally call for collaboration. When such situations are information-intensive, people often tend to use traditional tools for accessing and exchanging information that were not specifically designed for collaboration. This tendency is due to the tried, tested, and trusted nature of these tools. This confirms findings from previous studies, such as (Morris, 2008).

It was found that while the participants that were interviewed would like to get better systems designed for supporting collaborative projects, they may be unwilling to trade-off their traditional tools if the new systems do not fit well in their exiting practices. This has serious practical implications. While the importance of doing collaboration is well-accepted and the desire to have more sophisticated tools for CIS projects are theoretically prevalent, one may be hesitant to explore the possibilities of

collaborating with specially designed systems if such systems fail to incorporate the user's current practices and working environment.

In the interviews, the participants reported doing collaboration as a requirement for a project, out of a desire to distribute the work, or the appeal to bring diverse sets of skills to the process. In other words, the “to collaborate or not” dilemma was primarily addressed by either a pre-existing setup or the value of collaboration that the participants identified while working on a project. Their entry points for collaborative projects, i.e., how they start their collaborations, also varied from a weak social network to close colleagues. It is important that a CIS system facilitate all of these situations and inceptions of collaboration.

5.2 Design and development of Coagmento

Using the understanding from the interviews, the literature review, and good design principles, Coagmento, a system that allows multiple people to work together for CIS projects, was designed. The users can be remote or co-located, and work synchronously or asynchronously.

The earlier versions of Coagmento were designed to work on specific collections, and were used for several pilot runs and demonstrative walkthroughs. From those experiments and studies it was found that given an opportunity (or need), people can quickly learn a novel system such as Coagmento and appreciate the unique features it offers for CIS projects. Being restricted to using the specialized search interface that Coagmento supported and users' prior experience with tools such as integrated chat seemed to be the biggest obstacles for practical applications. The whole architecture of Coagmento was, therefore, redesigned and a more flexible and open system was created. Much of the design for this new Coagmento was driven by participatory design sessions conducted in Dr. Marchionini's HCI graduate seminar during the Fall 2008 and Fall

2009 semesters.

The result of all these studies, tests, and feedback was a system that allowed its users to use the web inside their familiar environment, viz., a web-browser, while offering support for collaboration. This was based on the suggestions by Grudin (1994) for developing an effective groupware system. From the literature (e.g., Rodden, 1991), it was clear that such a system needed to have three major components: *control*, *communication*, and *awareness*.

In Coagmento, the user has the full *control* over his personal usage and involvement with the system. At any time, the user can turn on or off Coagmento and resume his normal activities in the browser. Of course, in collaboration, a participant's control may depend on his role and responsibilities in that project. Coagmento, in this case, acts as a passive system, letting the collaborators work out the issue of control. Such a design was inspired by the works of Grudin (1994) as well as Denning and Yaholkovsky (2008), who proclaimed that it is solidarity, not software, that generates collaboration.

Coagmento allows the collaborators to exchange messages (*communication*) through a chat box in the browser's sidebar. Thus, one can communicate with his teammates working on the same project without leaving the browser, a familiar work environment for online information seeking. Coagmento also provides support for making annotations on a webpage, thus letting the collaborators exchange their views specific to a page they visited.

Finally, *awareness* in Coagmento is implemented in different ways. For one, the Coagmento toolbar constantly shows the project one is working on at the moment, and provides live statistics about that project. The Coagmento sidebar presents the team's history (bookmarks, queries, snippets) for the active project. Thus, one can be constantly aware of the status of the team and the project, without being distracted from the actual work, since Coagmento provides such information on the periphery

(even that can be turned off). Such way of providing awareness has been identified to be a highly valuable function of a CIS system (e.g., Bly et al., 1993).

5.3 Role of awareness in CIS environments

The new and improved version of Coagmento was then used for a large scale user study, involving 84 participants (42 pairs). The core issue that was addressed in this study was about the impact of different kinds of awareness support in CIS projects. Three conditions were created each with a different Coagmento interface. Each of them had the same support for control and communication, but the level of awareness varied between them.

With 42 pairs working on two different exploratory search tasks that spanned two sessions, it was found that having history information (bookmarks, queries used, snippets collected) is highly valuable for any information seeking task that spread across multiple sessions. In addition, while working in collaborative projects, it is very desirable to have access to similar history information about one's collaborators. In fact, those with only personal history support asked for the similar information about the group; they reported being unsatisfied and less engaged with the interface that was missing the group information awareness support. Design-wise, this information should be provided such that it does not distract the user from the actual task. This is referred to as peripheral awareness in CSCW literature (Liechti and Sumi, 2002).

It was discovered that, when group information is provided as a part of peripheral awareness, the collaborators could explore more information with the same amount of time and work. This is due to the fact that being constantly aware of the team's progress lets one focus on the task without having to worry about updating each other or repeating each other's work unintentionally. This supported the views of Gaver (1992), who argued for the provision of peripheral awareness that provides unobtrusive

information of the people, artifacts, and environment, while working on a collaborative project. More importantly, it was found that providing such awareness information does not add to the cognitive load, or lower user satisfaction. In fact, the provision of group peripheral awareness induced higher levels of engagement among the collaborators.

This study also demonstrated that, in an interactive, multi-session, and exploratory collaborative project, a single measure is not sufficient for evaluating the system or user's performance. For instance, it was not enough to see how many messages were exchanged between team members; we also had to obtain feedback on their engagement and cognitive load to understand if those messages were used to enhance the users' experience and productivity, or for coordination.

There are greater implications of this study beyond the importance and evaluation of awareness components as described here. Given that the study was done in a laboratory setting with limited amount of freedom to the participants, there was no need to coordinate certain events, such as user login, and task switching. If having proper awareness support was still found to be very crucial, one could imagine how real life situations may demand for even greater need of such awareness for effective coordination among the collaborators. This has implications on the design of such CIS system while balancing costs and benefits. For instance, in a field CIS system, we may need to add the support for traditional group awareness, which allows one to see the immediate status of his/her collaborators. This may introduce additional cost and concerns (e.g., privacy), but may provide benefits that could justify such cost.

Another way of understanding the importance of this study is looking back at Figure 3.4, where various CIS systems are traditionally organized on time-space dimensions. The placement of a CIS system on this figure has implications for its implementation, functionalities, and evaluation. For instance, Adobe Connect facilitates online meetings where the participants can share and discuss information. Such an environment will

fall under Synchronous-Remote collaboration in this figure. Thus, this environment needs to have a way to connect remote participants in real-time, and a shared space for exchanging and processing information. Coagmento is designed to provide support for collaborators working in synchronous or asynchronous mode, and primarily remotely located. The laboratory study discussed here simulated the situation of remotely located synchronous collaboration. Given that in real life, the participants may not be working synchronously with their teammates, they may require additional tools, such as a shared notepad (e.g., ShrEdit by Olson et al., 1993). During the study the participants were aware of their teammate being online and ready to respond. However, when there is no prerequisite of working in synchronous mode in collaboration, the participants asked for a way to know when their teammates were online. They also desired to have more active notification of chat messages and other critical actions, such as collecting relevant information. As we can see, while some of these components providing awareness were not required for this study (co-located, synchronous), they are desired in other situations.

The study was designed with a hypothesis that more awareness would help in making the collaborative experience more effective, engaging, and satisfactory for users. However, it was found that while the condition with the most awareness performed overall best, the condition that fell in between the highest awareness and the no awareness systems did the worst. This indicates the importance of providing an interface that matches the complexity and nature of the task at hand. Given that collaboratively seeking information in an interactive and multi-session setting is relatively more complex activity than an individual seeking information in a single session, condition 3 provided appropriate interface for the situation. Condition 1 incurred significantly more cost for coordinating events and kept the participants less engaged and aware, but the participants reported high level of satisfaction and ease of use. Condition 2,

on the other hand, had complex interface and did not provide enough added benefits as condition 3 did. Thus, it achieved low ratings for usability, while not offering any significant benefits over the baseline.

Overall, the results of this study suggest that designers should provide support for group awareness and other shared tools in a real life CIS systems. Although this study focused on synchronous collaboration, it is likely that group awareness support will be even more important when people search together asynchronously. The results suggest that the added complexity of group awareness can be implemented in such a way to avoid adding cognitive load to the users.

5.4 Limitations

The studies reported in this dissertation have several limitations that are important to point out here.

The first study involving personal interviews was limited in its scope since the interviewees were picked from information intensive domains. Had the participants been selected from other fields, it is possible that they would have very different kinds of collaborative experiences. The first study was also limited to a very small number of participants, and while there were common patterns found in their responses, a large scale study may be needed to develop a deeper understanding of people's motivations and methods for doing collaboration in information seeking projects.

The rest of the design studies were limited in their scopes as well, due to their dependence on the kind of task(s) as well as the functionalities provided by the experimental system. The pilot runs were also time-bound, not allowing the participants explore the given topic and engage with their teammates fully. The demonstrative walkthroughs and the participatory design sessions were limited in scope due to the background of the participants (almost all of them from information and library science field). A good

way to extend these studies is to involve participants from varying fields.

The laboratory study had participants from many different disciplines and varied backgrounds. There were, however, several limitations due to the design of the experiments, which are listed below.

1. The group size was only two. The dynamics among the collaborators is likely to change with a different size of group.
2. Both the participants in a given pair were required to have worked together on at least one collaborative project before. This is not true for all collaborations.
3. The topics for the tasks assigned to the participants were pre-defined. We may have observed a different level of engagement had the participants could choose their topics and define the tasks.
4. The participants were asked to collect relevant snippets from the web. They were told that the performance of a team (which can in turn earn them prizes) would be determined using three factors: (i) amount of information collected, (ii) quality of that information, and (iii) teamwork. Due to open-ended topics, artificial motivation, and the time-bound nature of the tasks, the participants focused primarily on collecting as many “good enough” snippets as possible. They may have tried to refrain from reflecting on or synthesizing information while completing this information seeking phase.
5. The assigned topics were also easily dividable. This encouraged almost all the teams to use divide-and-conquer as their primary strategy. Due to the time limit on tasks, the participants also tried to minimize their conversations. It is possible that if the tasks were non-dividable (e.g., brainstorming), the participants would have interacted more.

6. The study was designed to specifically understand the role and impact of awareness in CIS projects. While it provides an exploration of this particular issue of CIS systems, it is limited in scope to inform us about other issues with conformity. For instance, the communication channel for all the participants was facilitated using an integrated chat client; this study did not inform us the impact of other ways of providing communication.
7. By the very nature of this study (a laboratory experiment), it has limited applicability. The lessons learned from from this study should be considered with the knowledge that the real life situations can be significantly different than what was simulated in the laboratory. For instance, in reality, one may not have space and communication constraints that were presented in this study; in practice, one may be able to have face-to-face meetings with his/her collaborators and use other forms of communication media such as audio and video.
8. Some of the questions asked to the participants were suggestive in nature. For instance, those in condition 2 were asked if they would like to see group history instead of only personal history in their periphery, encouraging them to think that their interface was not good enough. This may have led to a biased ratings for their satisfaction with the system.
9. The majority of the analysis was done considering an individual as the unit, thus ignoring the group effect. A more comprehensive analysis is needed to understand group dynamics and comparing teams instead of individuals.

5.5 A framework for CIS environments

In this final section various lessons and experiences from this dissertation are synthesized to provide a framework for understanding user-centric CIS. This is done by

recognizing the general nature of collaboration on one hand and proposing a model for classifying various collaborative processes on the other hand.

5.5.1 A model of collaboration

In Section 2.1 (Figure 2.2) a model was proposed to explain collaboration in contrast to processes such as coordination and cooperation.

According to this model, a true collaboration requires all of the processes that it subsumes: communication, contribution, coordination, and cooperation. This model is collaboration centric, i.e., it is considering only those communication, contribution, coordination, and cooperation that are done to support collaboration.

More accurately, this model can be portrayed as shown in 5.1. Here, we can see that not all kinds of communication are intended for contribution, coordination, cooperation, or collaboration. Similarly a contribution may not always require some kind of communication and may not always need to be a part of coordination or cooperation. A coordination effort typically requires some kind of communication, but there are situations when this is not the case. One setting his wrist-watch according to his computer-watch does not require the exchange of any explicit messages.

Finally, a true collaboration requires some form of communication (due to its interactive nature), and contribution from the participants (due to its mutually beneficial nature). A collaborative project where the participants have intent and some goals, both coordination and cooperation are also essential. For instance, collaboration between a pilot and co-pilot requires them to not only coordinate their actions, but also cooperate with each other as the task of flying a plane requires.

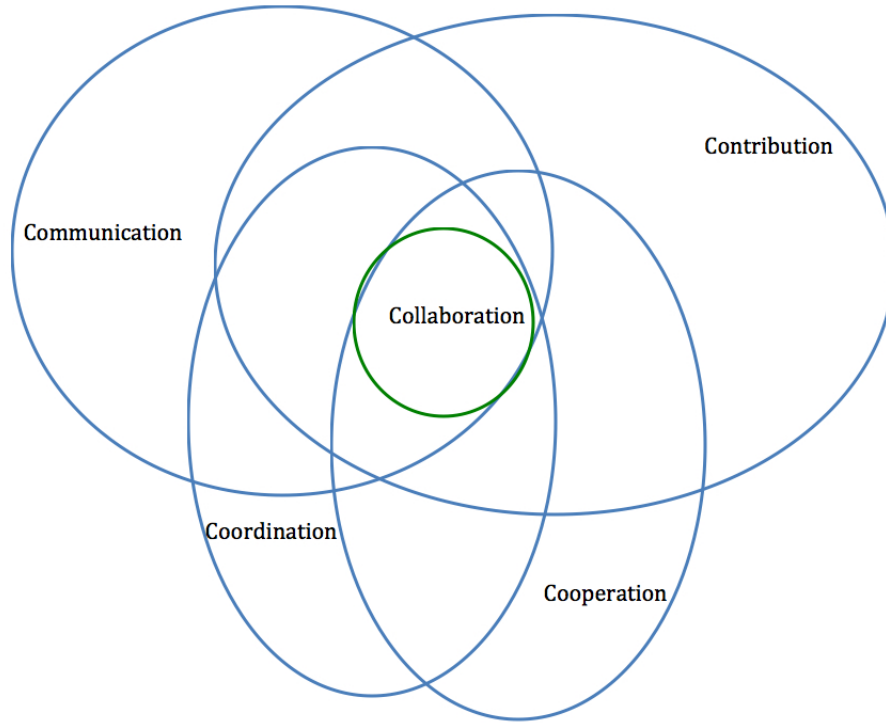


Figure 5.1: A modified set-based model of collaboration.

5.5.2 A model of CIS

Earlier in this dissertation CIS was put in the context of information seeking (IS) and collaboration as shown in Figure 1.1. It was argued that CIS encompasses some of the aspects of IR and/or IS, and it should be considered as a part of a larger collaboration.

Due to this configuration, it is imperative that we consider the context of the collaboration that a CIS process is a part of. Classically, CIS processes are laid out on time-space axes as shown in Figure 3.4. Golovchinsky, Pickens, and Back (2008) identified intent and depth of mediation as additional dimensions for classifying CIS systems. From the set of studies that were reported in this dissertation as well as previous works, a new framework is proposed that incorporates the following 11 dimensions for a CIS process. They are depicted in two parts in Figures 5.2 and 5.3, along with rationale behind incorporating these dimensions, and a few examples.

1. *Intent*

This dimension describes the level of intention one has in a collaborative process, or in other words, how explicitly collaboration is defined. This dimension is introduced to differentiate systems that are true collaborative to those that are merely collaborative filtering systems.

An intentional or explicitly defined collaboration is when various aspects of collaboration are clearly stated and understood. For instance, a group of students working on a science project together know that (i) they are collaborating, and (ii) who is responsible for doing what. When collaboration happens without explicit specifications, it can be considered unintentional or implicitly defined. For instance, visitors to Amazon.com receive recommendations based on other people's searching and buying behavior without knowing those people.

2. *Activeness*

The level of activeness is another important dimension in understanding the nature of a collaborative endeavor. This dimension is useful to talk about the involvement of a user in a group activity.

An active collaboration is similar to explicit collaboration with the key difference being the willingness and awareness of the user. For instance, when a user of Netflix rates a movie, he is actively playing a part in collaborating with other users. However, since he did not explicitly agree to collaborate with others; he may not even know those users. A passive collaboration is similar to implicit collaboration with the key difference being the willingness and awareness of the user. For instance, when a user visits a video on YouTube, he passively contributes to the popularity of that video, affecting the ranking and social relevance of that video for other users.

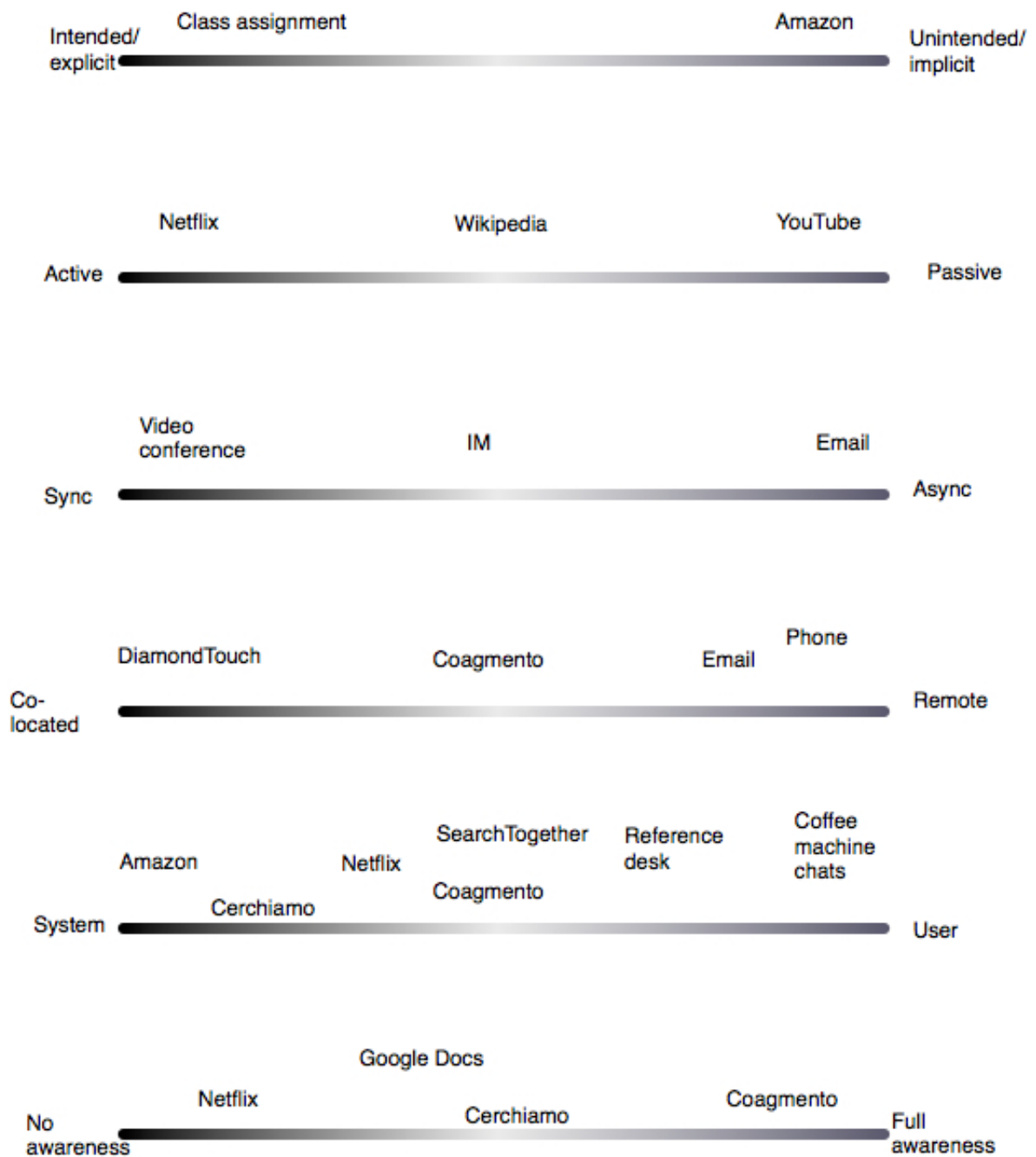


Figure 5.2: A framework for CIS environments (part 1 of 2).

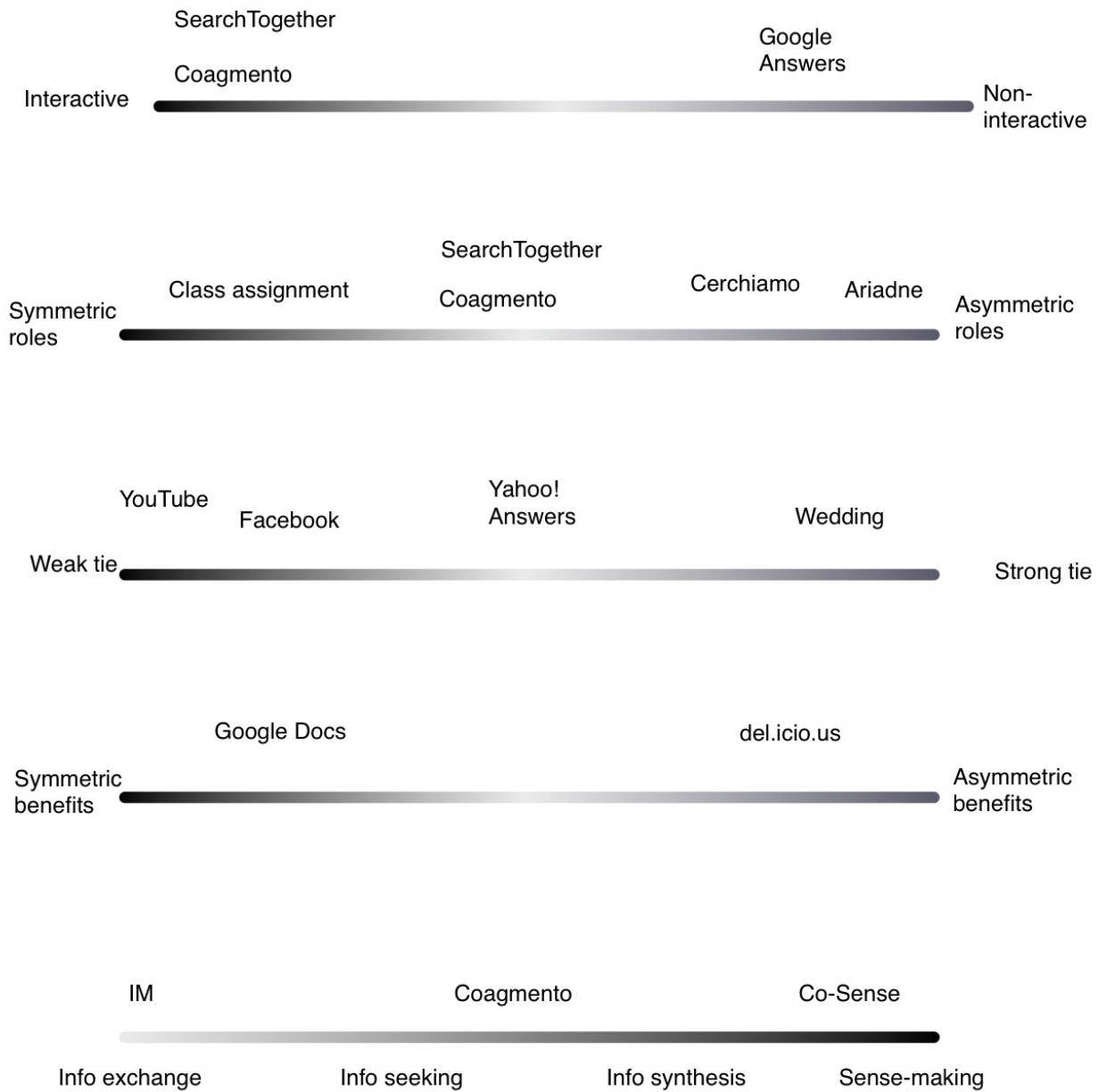


Figure 5.3: A framework for CIS environments (part 2 of 2).

The key difference between active and passive collaboration is the user’s willingness and control over the actions. In the case of active collaboration, the user agrees to do it (rating, comments), whereas in case of passive collaboration, the user has very little control (click-through, browsing patterns). To demonstrate this difference further, Table 5.1 lists some examples for these two dimensions.

Table 5.1: Differentiating between intent and activeness.

	Intended	Unintended
Active	Amazon wedding registry	Amazon book purchase
Passive	YouTube video viewing	Auto-subscribing

3. *Concurrency*

One of the traditional ways of classifying a collaborative process is by its concurrency. A video conference or a meeting typically requires the participants to be present in the same time (synchronous), whereas email could help a team work asynchronously. A chat program can support both synchronous and asynchronous collaborations, even though it is intended as a synchronous communication channel.

4. *Location*

This is another traditional dimension that is often used to place different collaborative systems in context. DiamondTouch (Smeaton et al., 2006b) requires the participants to be physically present around the system for a collaborative session. SearchTogether (Morris and Horvitz, 2007) and Coagmento, on the other hand, facilitate collaboration among remotely located participants.

5. *Role/mediation of system and user*

Collaboration can be entirely done by a group of people, and it could also incorporate support of systems, such as computers or phones. However, this dimension

says a bit more than that. In a collaborative project, the collaboration could be mediated by the system, in which case some underlying algorithm would drive the collaboration. Or the people or the users of the system could do mediation themselves, making the system (if it is used) a passive component. Cerchiamo (Golovchinsky, Adcock, Pickens, Qvarfordt, and Back, 2008a) is an example of the former, and Coagmento is an example of the latter.

6. *Level of awareness*

Gaver (1991) used awareness as a factor to identify different situations for collaboration. He claimed that less awareness is needed for division of labor, and that more casual awareness can lead to serendipitous communication, which can turn into collaboration. Awareness, therefore, is an important dimension to consider.

The amount and the kind of awareness provided in an environment with a group of people depends on several factors, including the cost and benefit of such awareness, available technology, and privacy. On one hand, services like Netflix and Amazon connect multiple users without hardly making them aware of one another. On Google Docs, one has workspace awareness, whereby one can work with the group's artifact in collaboration, but does not see his collaborators' individual contributions. Cerchiamo provides a system-driven collaboration, where the users have limited and filtered access to their collaborators' actions and results. Coagmento, on the other extreme, provides a very transparent interface, in which a user can be aware of the task at hand, the shared workspace, as well as the group's history and products.

7. *Level of interaction*

Once again, to differentiate systems with very little or no user interaction to those that are highly interactive, a dimension that considers level of user interaction is

needed. This dimension defines how interactive collaboration is. Systems such as SearchTogether and Coagmento are designed to support interactive collaboration. Google Answers, on the other hand, was a non-interactive service, where the information seeker could pose their questions to experts and receive answers without going back-and-forth.

8. *User roles*

While an effective collaboration must be *democratic* and *inclusive*, that is, it must be free of hierarchies of any kind and it must include all parties who have a stake in the problem (London, 1995), to include several scenarios of people working together, we should consider their roles in collaboration.

Division of labor and combining diverse sets of skills are two of the most attractive appeals for collaborative projects. Invariably, the former assumes symmetric roles, and the latter assumes asymmetric roles of the participants. For example, a group project for a class typically involves students who all have more or less the same background and skills. Ariadne (Twidale and Nichols, 1996), on the other hand, was designed to connect a patron to a reference librarian - each with a different background.

9. *Strength of the connection*

As realized during the literature review on social networking, a big difference between a social group and a collaborative group is the strength of the ties that connect them (Section 2.5). Often, one group could be transformed to the other one based on these ties. A collaborative endeavor could involve more or less of the social element.

Facebook is a social networking utility, where the users may not have stronger ties or common goals. Co-authoring a research article, on the other hand, involves

multiple parties being connected with a stronger bond.

10. *Balance of benefits*

This dimension follows the dimension of user roles. A typical collaboration is mutually beneficial for those participating. However, there can be a gradation of these benefits. Co-authoring a research article benefits all the involving authors, whereas one's collection of useful bookmarks on del.icio.us may benefit the author and the subscribers differently.

11. *Usage of information*

Finally, this dimension allows us to see how the information flows in the system. Often, information exchange is the focus of collaboration. An example is an online help service using chat. The other possible segments on this dimension are information seeking, information synthesis, and sense-making. A collaborative system could support one or several of these elements.

It is important to note here that all of these 11 dimensions are not independent. They have an interaction effect; i.e., fixing or altering of one dimension changes the rest appropriately. For instance, if we fix the 'Location' dimension to co-located, our options for 'Concurrency' dimension are reduced to synchronous, as the collaborators are likely to be meeting with each other at the same place and time. If, on the other hand, the collaborators could not meet face-to-face (remotely located), they may use synchronous (e.g., chat), or asynchronous (e.g., email) communication.

5.6 Future work

There are several interesting questions remained to be addressed in the field of CIS. The final section of this dissertation will highlight some of these questions, along with suggestions to extend the work reported here.

5.6.1 Extending the studies

As pointed out in the limitations section of this chapter, each of the studies reported in this dissertation was targeted toward a specific population - college students, professionals in social sciences, and knowledge workers. Further investigations are needed to study other specialized populations, such as families with health issues, and intelligent analysts working in teams. This may require to employ different methods of studying collaboration, including cognitive walkthroughs, and empirical observations.

The laboratory study can be extended to a field study, whereby allowing the participants to work without several limitations the controlled lab study had. The participants in a field study could work with the system as they please, creating their own projects of interest, and initiating collaborators with their colleagues and friends as they see fit. Running a field study over a long period of time (at least a few weeks) will also allow one to study long-term adoption effects, appropriation factor, and specialization with various features of the system.

As noted earlier in the limitations, one way to extend the reported lab study is by considering a team, rather than an individual, as the unit of analysis. Allowing any size of group for a collaborative project, and studying the group dynamics can be very complex procedure, but can also provide us with very insightful details into how people collaborate and what kind of support they need to make their collaborations more effective and engaging.

5.6.2 Theoretical issues

There are several interesting and important research questions that we need to address next, some of which are given below.

1. What additional tools are required to enhance existing methods of collaboration, given a specific domain? Such domains may include office environments, educational settings, or even domestic projects. In this dissertation, we saw that in order to extend an individual information seeking process to collaborative information seeking, we need to not only create a support system that connects the collaborators and makes it easy for them to communicate, but also provide appropriate and adequate awareness. Such requirements and specifications may vary from domain to domain.
2. How can we convert a social tie to a collaborative tie and vice versa? Through relevant literature review, this dissertation suggested that there may be ways to facilitate collaborative processes in social environments. Given the ubiquitous nature of online social networks, this issue merits further investigation.
3. What kind of visualization methods can be useful for CIS interfaces? Coagmento provided a very basic interface for viewing personal and shared information. However, the issue of having new kinds of dynamic and interactive interfaces appeared highly important during the participatory design sessions reported here. As it was noted, due to the narrow scope of this dissertation, this issue was kept out of further investigation, but deserves more advance treatment with interface designs and experimentations.
4. How to evaluate various aspects of collaborative information seeking, including system and user performance? It was pointed out in this dissertation that, no single measure is adequate for evaluation in CIS, and multiple measures need to be employed. It still remains a challenge to define and defend such measures for a given CIS situation.

5. How to measure the costs and benefits of collaboration? We saw that providing awareness induces additional cost, but if we are careful designing our CIS system, we can harvest many benefits of this awareness without causing any significant burden on the user. Similar investigations for other important issues for CIS, such as control and communication, are needed for developing a comprehensive understanding of various design issues in CIS.
6. What are the information seeking situations in which collaboration is beneficial? When does it not pay off? This dissertation started with an acknowledgement that collaboration is not always useful or desired, and then focused on only those situations where it is intentional and beneficial. Further investigations are needed to also study the situations where collaboration is potentially harmful. This understanding can help us do a better cost-benefit analysis of doing collaboration in a given situation.
7. How to extend individual information seeking, synthesis, and sense-making models to incorporate collaboration? This dissertation was focused on people working on information seeking process in collaboration. It is possible that certain form of information synthesis and sense-making are also taking place during such information seeking process. For the sake of simplifying the studies, such possibilities were ignored here, and needs further investigation in the future studies. The current chapter proposed a new framework for CIS. Similar frameworks for collaborative information synthesis and sense-making are needed.

This dissertation has carved a new way for thinking about user-centric collaborative information seeking, and added notable contribution to our knowledge of awareness in CIS. This augmented understanding and the framework presented here will, hopefully, help us in further explorations in the field of CIS.

Bibliography

- Adomavicius, G. and Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749.
- Almack, J. C. (1922). The influence of intelligence on the selection of associates. *School and Society*, 16:529–530.
- Amershi, S. and Morris, M. R. (2008). CoSearch: a system for co-located collaborative Web search. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 1647–1656, Florence, Italy.
- Aneiros, M. and Estivill-Castro, V. (2003). Foundations of unconstrained collaborative web browsing with awareness. In *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, pages 18–25, Beijing, China.
- Aneiros, M. and Estivill-Castro, V. (2005). Usability of real-time unconstrained WWW-co-browsing for educational settings. In *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, pages 105–111, Compiegne University of Technology, France.
- Anick, P. (2003). Using terminological feedback for web search refinement - a log-based study. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 88–95.
- Aristotle (384 BC – 322 BCa). Ethics.
- Aristotle (384 BC – 322 BCb). Metaphysica. 10f-1045a.
- Aristotle (384 BC – 322 BCc). Politics I.2.
- Aslam, J. A. and Montague, M. (2001). Models for metasearch. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 276–284.
- Austin, A. E. and Baldwin, R. G. (1991). *Faculty Collaboration: Enhancing the Quality of Scholarship and Teaching*. J-B ASHE Higher Education Report Series (AEHE). Jossey-Bass.
- Baecker, R. M. (1995). *Readings in Human-Computer Interaction: Towards the Year 2000*. Morgan Kaufmann.

- Baeza-Yates, R. and Pino, J. A. (1997). A first step to formally evaluate collaborative work. In *Proceedings of GROUP*, pages 56–60.
- Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science*, 30(4):205–214.
- Bates, M. J. (2002). The cascade of interactions in the digital library interface. *Information Processing and Management*, 38(3):381–400.
- Belkin, N., Marchetti, P., and Cool, C. (1993). Braque: Design of an interface to support user interaction in information retrieval. *Information Processing and Management*, 29(3):325–344.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5:133–143.
- Belkin, N. J., Cool, C., Stein, A., and Thiel, U. (1995). Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems with Applications*, 9:379–395.
- Belkin, N. J. and Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):29–38.
- Berelson, Lazarsfeld, and McPhee (1954). *Voting*. University of Chicago Press, Chicago, IL.
- Billsus, D., Brunk, C. A., Evans, C., Gladish, B., and Pazzani, M. (2002). Adaptive interfaces for ubiquitous web access. *Communications of the ACM*, 45(5):34–38.
- Blackwell, A. F., Stringer, M., Toye, E. F., and Rode, J. A. (2004). Tangible interface for collaborative information retrieval. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 1473–1476, Vienna, Austria. ACM Press.
- Blake, C. and Pratt, W. (2006). Collaborative information synthesis i: a model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology*, 57(13):1740–1749.
- Bly, S. A., Harrison, S. R., and Irwin, S. (1993). Media spaces: bringing people together in a video, audio, and computing environment. *Communications of the ACM*, 36(1):28–46.
- Bott, H. (1928). Observation of play activities in a nursery school. *Genetic Psychology Monographs*, 4:44–48.
- Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the Seventh World Wide Web Conference*.

- Bruce, H., Fidel, R., Pejtersen, A. M., Dumais, S. T., Grudin, J., and Poltrock, S. (2003). A comparison of the collaborative information retrieval behaviour of two design teams. *The New Review of Information Behaviour Research*, pages 139–153.
- Buckley, C., Salton, G., and Allen, J. (1994). The effect of adding relevance information in a relevance feedback environment. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 292–300, New York, NY. Springer-Verlag.
- Cabri, G., Leonardi, L., and Zambonelli, F. (1999). Supporting cooperative WWW browsing: a proxy-based approach. Technical report, University of Modena, Italy.
- Callan, J. (2002). *Distributed Information Retrieval*, volume 7 of *The Information Retrieval Series*, pages 127–150. Springer Netherlands.
- Chalmers, M. (2002). Awareness, representation and interpretation. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):389–409.
- Chrislip, D. D. and Larson, C. E. (1994). *Collaborative Leadership: How Citizens and Civic Leaders Can Make a Difference*. Jossey-Bass, San Francisco.
- d. m. boyd and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1).
- Denning, P. J. (2007). Mastering the mess. *Communications of the ACM*, 50(4):21–25.
- Denning, P. J. and Yaholkovsky, P. (2008). Getting to “We”. *Communications of the ACM*, 51(4):19–24.
- Dervin, B. (1977). Useful theory for librarianship: Communication, not information. *Drexel Library Quarterly*, 13(3):16–32.
- Dervin, B. and Nilan, M. S. (1986). Information needs and uses. *Annual Review of Information Science and Technology (ARIST)*, 21:3–33.
- Desheids, O. and Kara, A. (2000). Homophily effects in advertising communication moderated by spokesperson’s credibility. In *Proceedings of Marketing in a Global Economy*, pages 315–325.
- Donath, J. S. and Robertson, N. (1994). The sociable web. In *Proceedings of the World Wide Web (WWW) Conference*, CERN, Geneva, Switzerland.
- Dourish, P. and Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW)*, pages 107–114, Toronto, Ontario.
- Dourish, P. and Bly, S. A. (1992). Portholes: supporting awareness in a distributed work group. In *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW)*, pages 541–547, Toronto, Canada. New York, N.Y.: ACM Press.

- Dumais, S. T., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*. ACM Press.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, 46:318–338.
- Ellis, D., Cox, D., and Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49:356–369.
- Ellis, D. and Haugan, M. (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53(4):384–403.
- Erickson, T. (2006). From PIM to GIM: Personal Information Management in Group Contexts. *Communications of the ACM*, 49(1):40–43.
- Esenther, A. W. (2002). Instant co-browsing: Lightweight real-time collaborative web browsing. In *Proceedings of the World Wide Web (WWW) Conference*, pages 107–114, Honolulu, Hawaii, USA.
- Evans, B. M. and Chi, E. H. (2008). Towards a model of understanding social search. In *Proceedings of JCDL 2008 Workshop on Collaborative Exploratory Search*, Pittsburgh, PA.
- Farooq, U., Ganoë, C. H., Carroll, J. M., and Giles, C. L. (2009). Designing for e-science: Requirements gathering for collaboration in CiteSeer. *International Journal of Human Computer Studies*, 67:297–312.
- Fidel, R., Bruce, H., Dumais, S. T., Grudin, J., Poltrock, S., and Pejtersen, A. M. (1999). Collaborative information retrieval. Technical report, University of Washington.
- Fidel, R., Bruce, H., Pejtersen, A. M., Dumais, S. T., Grudin, J., and Poltrock, S. (2000a). Collaborative Information Retrieval (CIR). *The New Review of Information Behaviour Research*, pages 235–247.
- Fidel, R., Bruce, H., Pejtersen, A. M., Dumais, S. T., Grudin, J., and Poltrock, S. (2000b). Collaborative information retrieval (cir). In *ISIC*.
- Fidel, R., Pejtersen, A. M., Cleal, B., and Bruce, H. (2004). A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval. *Journal of the American Society for Information Science and Technology*, 55(11):939–953.

- Flora, C. B., Flora, J. L., and Fey, S. (2004). *Rural Communities: Legacy and Change*. Westview Press.
- Foster, J. (2006). Collaborative information seeking and retrieval. *Annual Review of Information Science and Technology (ARIST)*, 40:329–356.
- France-Presse, A. (2007). MySpace exposes sex predators. <http://www.news.com.au/heraldsun/story/0,21985,21775032-11869,00.html>.
- Freeman, L. C. (1996). Some antecedents of social network analysis. *Connections*, 19(1):39–42.
- Freyne, J., Smyth, B., Coyle, M., Balfe, E., and Briggs, P. (2004). Further experiments on collaborative ranking in community-based web search. *Artificial Intelligence Review*, 21:229–252.
- Fu, X., Kelly, D., and Shah, C. (2007). Using collaborative queries to improve retrieval for difficult topics. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 879–880, Amsterdam, The Netherlands.
- Furnas, G. W. (1985). Experience with an adaptive indexing scheme. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 131–135.
- Gaver, W. W. (1991). Sound support for collaboration. In Bannon, L. J., Robinson, M., and Schmidt, K., editors, *Proceedings of European Conference on Computer Supported Cooperative Work (ECSCW)*, Amsterdam, The Netherlands. Dordrecht: Kluwer Academic Publishers.
- Gaver, W. W. (1992). The affordances of media spaces for collaboration. In Turner, J. and Kraut, R., editors, *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW)*, pages 17–24, Toronto, Canada.
- Gaver, W. W., Moran, T. P., MacLean, A., Löfstrand, L., Dourish, P., Carter, K., and Buxton, W. A. (1992). Realizing a video environment: EuroPARC’s RAVE system. In Bauersfeld, P., Bennett, J., and Lynch, G., editors, *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 27–35, Monterey, California. New York, N.Y.: ACM Press.
- Gerosa, L., Giordani, A., Ronchetti, M., Soller, A., and Stevens, R. (2004). Symmetric synchronous collaborative navigation. In *Proceedings of the 2004 IADIS International WWW/Internet Conference*, pages 1–7, Madrid, Spain.
- Ghani, J. A., Supnick, R., and Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. In *Proceedings of International Conference on Information Systems*, pages 229–237, Minneapolis, MN, USA. University of Minnesota.

- Golovchinsky, G., Adcock, J., Pickens, J., Qvarfordt, P., and Back, M. (2008a). Cerchiamo: a collaborative exploratory search tool. In *Proceedings of Computer Supported Cooperative Work (CSCW)*.
- Golovchinsky, G., Pickens, J., and Back, M. (2008b). A taxonomy of collaboration in online information seeking. In *Proceedings of JCDL 2008 Workshop on Collaborative Exploratory Search*, Pittsburgh, PA.
- Govern, J. M. and Marsch, L. A. (2001). Development and Validation of the Situational Self-Awareness Scale. *Consciousness and Cognition*, 10(3):366–378.
- Graham, T. C. N. (1997). GroupScape: Integrating Synchronous Groupware and the World Wide Web. In *Proceedings of INTERACT*, Sydney, Australia. Chapman and Hall.
- Gray, B. (1989). *Collaborating: Finding Common Ground for Multiparty Problems*. Jossey-Bass.
- Greenberg, S. and Roseman, M. (1996). GroupWeb: A WWW Browser as Real Time Groupware. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 271–272, Boston, MA. ACM Press.
- Grudin, J. (1994). Groupware and social dynamics: eight challenges for developers. *Communications of the ACM*, 37(1):92–105.
- Gutwin, C. and Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446.
- Hagman, E. P. (1933). The championships of preschool children. *University of Iowa Studies in Child Welfare*, 7:10–69.
- Han, R., Perret, V., and Naghshineh, M. (2000). WebSplitter: A Unified XML Framework for Multi-Device Collaborative Web Browsing. In *Proceedings of Computer Supported Cooperative Work (CSCW)*, pages 221–230. ACM Press.
- Hansen, P. and Jarvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing and Management*, 41:1101–1119.
- Harper, R. H., Hughes, J. A., and Shapiro, D. Z. (1989a). The functionality of flight strips in ATC work. The report for the Civil Aviation Authority. Technical report, Lancaster Sociotechnics Group, Department of Sociology, Lancaster University.
- Harper, R. H., Hughes, J. A., and Shapiro, D. Z. (1989b). Working in harmony: An examination of computer technology in air traffic control. In *Proceedings of*

- European Conference on Computer Supported Cooperative Work (ECSCW)*, pages 72–86, Gatwick, London.
- Heath, C., Svensson, M. S., Hindmarsh, J., Luff, P., and vom Lehn, D. (2002). Configuring awareness. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):317–347.
- Heath, C. C. and Luff, P. (1991). Collaborative activity and technological design: Task coordination in london underground control rooms. In Bannon, L. J., Robinson, M., and Schmidt, K., editors, *Proceedings of European Conference on Computer Supported Cooperative Work (ECSCW)*, pages 65–80, Amsterdam, The Netherlands. Dordrecht: Kluwer Academic Publishers.
- Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing and Management*, 44:957–962.
- Hubbard, R. M. (1929). A method of studying spontaneous group formation. In Thomas, D. S., editor, *Some New Techniques for Studying Social Behavior*, pages 76–85. Teachers College, Columbia University, Child Development Monographs, New York, NY.
- Hust, A. (2005). Query expansion methods for collaborative information retrieval. *Informatik*, 19(4):224–238.
- Hust, A., Junker, M., and Dengel, A. (2004). A mathematical model for improving retrieval performance in collaborative information retrieval. *Kluwer Information Retrieval Special Issue: Advances in Mathematical/Formal Methods in Information Retrieval*, pages 1–28.
- Hust, A., Klink, S., Junker, M., and Dengel, A. (2002). Query reformulation in collaborative information retrieval. In *Information and Knowledge Sharing*.
- Hyldegard, J. (2006). Collaborative information behaviour - exploring Kuhlthau’s Information Search Process model in a group-based educational setting. *Information Processing and Management*, 42:276–298.
- Hyldegard, J. (2009). Beyond the search process - exploring group members’ information behavior in context. *Information Processing and Management*, 45:142–158.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52:3–50.
- Karamuftuoglu, M. (1998). Collaborative Information Retrieval: Toward a Social Informatics View of IR Interaction. *Journal of the American Society for Information Science*, 49(12):1070–1080.
- Kautz, H., Selman, B., and Shah, M. (1997). Referral web: Combining social networks and collaborative filtering. *Communications of the ACM*, 40(3):63–65.

- Keller, R. M., Wolfe, S. R., Chen, J. R., Rabinowitz, J. L., and Mathe, N. (1997). A bookmarking service for organizing and sharing URLs. *Computer Networks and ISDN Systems*, 29:1103–1114.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632.
- Klink, S. (2001). Query reformulation with collaborative concept-based expansion. In *First International Workshop on Web Document Analysis*, pages 19–22.
- Krebs, V. (2006). Social network analysis, a brief introduction. <http://www.orgnet.com/sna.html>.
- Krishnappa, R. (2005). Multi-User Search Engine: Supporting Collaborative Information Seeking and Retrieval. Master’s thesis, University of Missouri-Rolla.
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user’s perspective. *Journal of the American Society for Information Science and Technology*, 42(5):361–371.
- Kuhlthau, C. C. (1994). *Seeking meaning: a process approach to library and information services*. Ablex Publishing, Norwood, NJ.
- Kuhlthau, C. C. (2005). Towards collaboration between information seeking and information retrieval. *Information Research*, 10(2).
- Large, A., Beheshti, J., and Rahman, T. (2002). Gender differences in collaborative web searching behavior: an elementary school study. *Information Processing and Management*, 38:427–433.
- Laurillau, Y. (1999). Synchronous Collaborative Navigation on the WWW. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 308–309.
- Laurillau, Y. and Nigay, L. (2002). CoVitesse: A Groupware Interface for Collaborative Navigation on the WWW. In *Proceedings of the ACM conference on Computer Supported Cooperative Work (CSCW)*, pages 236–240, New Orleans, Louisiana, USA.
- Lauwers, J. C. and Lantz, K. A. (1990). Collaboration awareness in support of collaboration transparency: requirements for the next generation of shared window systems. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 303–311. New York, N.Y.: ACM Press.
- Lazarsfeld, P. F. and Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. In Berger, M., Abel, T., and Page, C. H., editors,

- Freedom and Control in Modern Society*, pages 18–66. Van Nostrand, New York, NY.
- Lee, R. B. and DeVore, I., editors (1968). *Man the Hunter*. Aldine Publishing Company.
- Levy, D. M. and Marshall, C. C. (1994). What color was George Washington’s white horse? A look at assumptions unerlying digital libraries. In *Proceedings of Digital Libraries*, pages 163–169.
- Lewis, J. R. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human-Computer Interaction*, 7(1):57–58.
- Liebel, G. (2008). Why Users Love Orkut.
<http://usability.about.com/od/websiteaudiences/a/Orkut.htm>.
- Liechti, O. and Sumi, Y. (2002). Editorial: Awareness and the WWW. *International Journal of Human Computer Studies*, 56(1):1–5.
- Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80.
- London, S. (1995). Collaboration and community.
<http://scottlondon.com/reports/ppcc.html>.
- Lucas, G. (1973). American Graffiti. Movie.
- Luo, J. S. (2007). Social Networking: Now Professionally Ready. *Primary Psychiatry*, 14(2):21–24.
- Maderazo, J. W. (2007). Orkut, Friendster Get Second Chance Overseas.
http://www.pbs.org/mediashift/2007/06/try_try_againorkut_friendster.html.
- Maekawa, T., Hara, T., and Nishio, S. (2006). A collaborative web browsing system for multiple mobile users. In *Proceedings of IEEE Conference on Pervasive Computing and Communications (PERCOM)*.
- Malone, T. W. (1988). What is coordination theory? Technical Report SSM WP # 2051-88, Massachusetts Institute of Technology, Boston, MA.
- Malone, T. W., Grant, K. R., Turbak, F. A., Brobst, S. S., and Cohen, M. D. (1987). Intelligent information sharing systems. *Communications of the ACM*, 30(5):390–402.
- Malone, T. W. and Lai, K. (1988). Object Lens: a spreadsheet for cooperative work. In *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW)*, Portland, OR.

- Mantei, M. M., Baecker, R. M., Sellen, A. J., Buxton, W. A. S., and Milligan, T. (1991). Experiences in the use of a media space. In Robertson, S., Olson, G., and Olson, J., editors, *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 203–208, New Orleans, Louisiana. New York, N.Y.: ACM Press.
- Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, 40(1):54–66.
- Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge University Press, Cambridge, England.
- Menkov, V., Neu, D. J., and Shi, Q. (2000). AntWord: A Collaborative Web Search Tool. In *Workshop on Distributed Communities on the Web (DCW)*, Quebec City, Quebec, Canada.
- Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., and Riedl, J. (2003). MovieLens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of International Conference on Intelligent User Interfaces*, pages 263–266, Miami, Florida, USA.
- Miller, S. M. and Mangan, C. E. (1983). Interesting effects of information and coping style in adapting to gynaecological stress: should a doctor tell all? *Journal of Personality and Social Psychology*, 45:223–226.
- Moreno, J. L. (1934). *Who Shall Survive?* Nervous and Mental Disease Publishing Company, Washington, D.C.
- Morris, M. R. (2007a). Collaborating alone and together: Investigating persistent and multi-user web search activities. Technical Report MSR-TR-2007-11, Microsoft.
- Morris, M. R. (2007b). Interfaces for collaborative exploratory web search: Motivations and directions for multi-user design. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems 2007 Workshop on Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction*, pages 9–12.
- Morris, M. R. (2008). A survey of collaborative web search practices. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems Conference on Human Factors in Computing Systems*, pages 1657–1660, Florence, Italy.
- Morris, M. R. and Horvitz, E. (2007a). S^3 : Storable, Shareable, Search. In *Interact*.
- Morris, M. R. and Horvitz, E. (2007b). SearchTogether: An Interface for Collaborative Web Search. In *ACM Symposium on User Interface Software and Technology (UIST)*, pages 3–12, Newport, RI.

- Morris, M. R., Paepcke, A., and Winograd, T. (2006). TeamSearch: Comparing Techniques for Co-Present Collaborative Search of Digital Media. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)*, pages 97–104, Adelaide, South Australia.
- Morris, M. R., Teevan, J., and Bush, S. (2008). Enhancing Collaborative Web Search with Personalization: Groupization, Smart Splitting, and Group Hit-Highlighting. In *Proceedings of Computer Supported Cooperative Work (CSCW)*, San Diego, CA.
- Mu, X. (2004). *Smartlinks in a video-based collaborative distance learning system: a cognitive model and evaluation study*. PhD thesis, UNC Chapel Hill.
- Newcomb, A. F. and Bagwell, C. L. (1995). Children’s friendship relations - a meta-analytic review. *Psychological Bulletin*, 117:306–347.
- O’Hear, S. (2007). Bebo overtakes MySpace in the UK. <http://blogs.zdnet.com/social/?p=279>.
- Olson, G. M., Olson, J. S., Carter, M. R., and Storrosten, M. (1992). Small group design meetings: An analysis of collaboration. *Human-Computer Interaction*, 7(4):347–374.
- Olson, J. S., Olson, G. M., Storrorsten, M., and Carter, M. (1993). Groupwork close up: a comparison of the group design process with and without a simple group editor. *ACM Transactions on Information Systems*, 11(4):321–348.
- Osborne, D. and Gaebler, T. (1992). *Reinventing Government*. Addison-Wesley Publishing Company.
- Pickens, J. and Golovchinsky, G. (2007). Collaborative Exploratory Search. In *Proceedings of Workshop on Human-Computer Interaction and Information Retrieval*, pages 21–22, MIT CSAIL, Cambridge, Massachusetts, USA.
- Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. (2008). Algorithmic mediation for collaborative exploratory search. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, Singapore.
- Plato (427 BC – 347 BC). Phaedrus.
- Poltrock, S., Grudin, J., Dumais, S. T., Fidel, R., Bruce, H., and Pejtersen, A. M. (2003). Information seeking and sharing in design teams. In *GROUP*, pages 239–247.
- Portnoy, F. (1986). Collaboration: Go for it. *Foundation News*, pages 59–61.
- Prekop, P. (2002). A qualitative study of collaborative information seeking. *Journal of Documentation*, 58(5):538–547.

- Reddy, M. C. 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):256–273.
- Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):56–58.
- Roberts, N. C. and Bradley, R. T. (1991). Stakeholder collaboration and innovation: a study of public policy initiation at the state level. *The Journal of Applied Behavioral Science*, 27(2):209.
- Rodden, T. (1991). A Survey of CSCW Systems. *Interacting with Computers*, 3(3):319–353.
- Rodden, T. (1996). Populating the application: a model of awareness for cooperative applications. In *Proceedings of Computer Supported Cooperative Work (CSCW)*, pages 87–96, Boston, MA. New York, N.Y.: ACM Press.
- Rodden, T. and Blair, G. (1991). CSCW and distributed systems: the problem of control. In *Proceedings of ECSCW*, pages 49–64, Amsterdam, The Netherlands.
- Romano Jr., N. C., Roussinov, D., Jay F. Nunamaker, J., and Chen, H. (1999). Collaborative information retrieval environment: Integration of information retrieval with group support systems. In *Proceedings of the 32nd Hawaii International Conference on System Sciences*, pages 1–10.
- Root, R. W. (1988). Design of a multi-media vehicle for social browsing. In *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW)*, pages 25–38.
- Roxborough, S. and Masters, C. (2007). Social nets engage in global struggle. http://www.adweek.com/aw/national/article_display.jsp?vnu_content_id=1003653287.
- Rutgers (2009). Rutgers DISCIPLE Project.
- Sandor, O., Bogdan, C., and Bowers, J. (1997). Aether: An awareness engine for CSCW. In Hughes, J., Prinz, W., Rodden, T., and Schmidt, K., editors, *Proceedings of European Conference on Computer Supported Cooperative Work (ECSCW)*, pages 221–236, Lancaster, U.K. Dordrecht: Kluwer Academic Publishers.
- Saracevic, T. (1996). Modeling interaction in information retrieval (IR): a review and proposal. In *Annual Meeting of the American Society for Information Science*, pages 3–9, Silver Spring, MD.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: extension and applications. In *Proceedings of American Society of Information Science*, pages 313–327.

- Schmidt, K. (2002). The Problem with ‘Awareness’: Introductory Remarks on ‘Awareness in CSCW’. *Computer Supported Cooperative Work (CSCW)*, 11(3):285–298.
- Shah, C. (2008). Toward Collaborative Information Seeking (CIS). In *Proceedings of JCSDL 2008 Workshop on Collaborative Exploratory Search*, Pittsburgh, PA.
- Shah, C. (2010). Working in collaboration - what, why, and how? In *Proceedings of Collaborative Information Retrieval workshop at CSCW 2010*, Savannah, GA.
- Simone, C. and Bandini, S. (2002). Integrating awareness in cooperative applications through the reaction-diffusion metaphor. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):495–530.
- Smeaton, A. F., Foley, C., Gurrin, C., Lee, H., and Givney, S. M. (2006a). 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*, pages 149–156, Adelaide, Australia.
- Smeaton, A. F., Lee, H., Foley, C., and Givney, S. M. (2006b). Collaborative video searching on a tabletop. *Multimedia Systems Journal*, 12(4):375–391.
- Smeaton, A. F., Lee, H., Foley, C., Givney, S. M., and Gurrin, C. (2006c). Físchlár-DiamondTouch: Collaborative Video Searching on a Table. In *SPIE Electronic Imaging - Multimedia Content Analysis, Management, and Retrieval*, volume 6073, San Jose, CA.
- Smyth, B., Balfe, E., Boydell, O., Bradley, K., Briggs, P., Coyle, M., and Freyne, J. (2005). A live-user evaluation of collaborative web search. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Edinburgh, Scotland.
- Smyth, B., Balfe, E., Briggs, P., Coyle, M., and Freyne, J. (2003). Collaborative web search. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1417–1419, Acapulco, Mexico. Morgan Kaufmann.
- Sonnenwald, D. H. (1996). Communication roles that support collaboration during the design process. *Design Studies*, 17(3):277–301.
- Spink, A., Greisdorf, H., and Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management*, 34(5):599–621.
- Straus, D. and Williams, D. (1986). *Collaborative Problem Solving in Local Agenda Setting Processes*. Lincoln Institute of Land Policy, Cambridge, MA.
- Sullivan, D. (2005). Search engine sizes.
<http://searchenginewatch.com/showPage.html?page=2156481>.

- Surowiecki, J. (2004). *Wisdom of Crowds : Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday Publishing.
- Talja, S. and Hansen, P. (2006). *Information sharing*. New Directions in Human Information Behavior. Springer.
- Taylor-Powell, E., Rossing, B., and Geran, J. (1998). Evaluating collaboratives: Reaching the potential. Technical report, University of Wisconsin-Extension, Madison, Wisconsin.
- Teevan, J., Dumais, S. T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, pages 449–456.
- Theobald, R. (1987). *The rapids of change*. Knowledge Systems, Inc.
- Twidale, M. and Marty, P. (2005). Exploring the connections between abstraction and appropriation. *International reports on socio-informatics*, 2(2):43–47.
- Twidale, M. B. and Nichols, D. M. (1996). Collaborative browsing and visualisation of the search process. In *Proceedings of Aslib*, volume 48, pages 177–182.
- Twidale, M. B., Nichols, D. M., and Paice, C. D. (1995). Supporting collaborative learning during information searching. In *Proceedings of Computer Supported Collaborative Learning (CSCL)*, pages 367–374, Bloomington, Indiana.
- Twidale, M. B. T., Nichols, D. M. N., and Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing and Management*, 33(6):761–783.
- University of Washington (2007). Collaborative information retrieval. Available from <http://projects.ischool.washington.edu/cir/>.
- Wasserman, S. and Faust, K. (1994). *Social Networks Analysis: Methods and Applications*. Cambridge University Press, Cambridge, MA.
- Wellman, B. (1926). The school child’s choice of companions. *Journal of Educational Research*, 14:126–132.
- Westbrook, L. (1993). User needs: a synthesis and analysis of current theories for the practitioner. *RQ*, 32:541–549.
- Wikipedia (2008). Cloud computing. http://en.wikipedia.org/wiki/Cloud_computing.
- Wilbur, S. B. and Young, R. E. (1988). The COSMOS project: a multi-disciplinary approach to design of computer supported group working. In Speth, R., editor, *EUTECO 88: Research into Networks and Distributed Applications*, Vienna, Austria.

- Wilson, M. L. and m.c. schraefel (2008). Evaluating collaborative search interfaces with information seeking theory. In *Workshop on Collaborative Information Retrieval*, Pittsburgh, PA.
- Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37(1):3–15.
- Wilson, T. D. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing and Management*, 33(4):551–572.
- Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55(3):249–270.
- Wittenburg, K., Das, D., Hill, W., and Stead, L. (1995). Group Asynchronous Browsing on the World Wide Web. In *Proceedings of the World Wide Web (WWW) Conference*, Boston, MA.

Appendix A

Definitions

This appendix presents definitions for some of the terms used in this article. Instead of universal meanings, these definitions are rather my interpretations of the terms for their specific uses for my work.

- *Information need*: This involves fact finding, exploration of a topic, content consumption (e.g., read a document, view a video, buy a product), negotiations (e.g., auctions), etc.
- *Explicit collaboration*: When various aspects of collaboration are clearly stated and understood. For instance, a group of students working on a science project together know that (i) they are collaborating, (ii) who is responsible for doing what.
- *Implicit collaboration*: When collaboration happens without explicit specifications. For instance, visitors to Amazon.com receive recommendations based on other people's searching and buying behavior without knowing those people.
- *Active collaboration*: This is similar to explicit collaboration with the key difference being the willingness and awareness of the user. For instance, when a user of Netflix rates a movie, he is actively playing a part in collaborating with other users. However, since he did not explicitly agree to collaborate with others; he may not even know those users.
- *Passive collaboration*: This is similar to implicit collaboration with the key difference being the willingness and awareness of the user. For instance, when a user visits a video on YouTube, he passively contributes to the popularity of that video, affecting the ranking and popularity of that video for others. The key difference between active and passive collaboration is user's willingness and control over the actions. In case of active collaboration, user agrees to do it (rating, comments), whereas in case of passive collaboration, user has very little control (click-through, browsing patterns).
- *timeframe*: A short slot of time. Two events happening in the same timeframe do not need to be happening exactly at the same time, but they do take place not too far from each other in temporal context. For instance, this morning's online discussions on a breaking news story is considered in the same timeframe. If one is commenting on a news story from the last week, that is out of that news story's timeframe.

- *Personalization*: System configuration for a given user based on his profile, preferences and/or behavioral patterns.
- *Groupization*: System configuration for a given group based on its profile, preferences and/or behavioral patterns.
- *Recommendation*: System configuration for a given user based on his matched profile, preferences and/or behavioral patterns with other users in the network.
- *Agent*: An entity - user or system - in a collaborative environment. A computer network is made by connecting agents that are computers. A focus group consists of agents that are humans.
- *Users*: Refers to the humans. In some cases, they may be using a system, but not be a part of a collaborative.
- *Participants*: Refers to the humans that are parts of a collaborative. In some cases, they may not even use a system.
- *System*: Refers to machines or automated mechanism.
- *Environment*: A set of objects and attributes that may include users, systems, and their context.
- *Symmetric role*: When all the users in the group are given the same powers and responsibilities, they are playing symmetric role.
- *Asymmetric role*: When each user in the group has different powers and/or responsibilities, they are said to be playing asymmetric roles.

Appendix B

Study-1 Interview Questionnaire

1. How often do you search on the Internet or library catalogues?
2. Does it ever happen that you wish someone could help you with your search?
If “yes”:
 - (a) Did you ever receive help in such a situation? Please describe how this happened (e.g., how was it initiated, how did it proceed, what was the outcome)? Were there specific methods or tools that were used? Do you think such help must be given by a human or could a machine have helped? For which functions might an automated system have helped?
 - (b) For times when you wished you had some help with searching but did not receive any help, what kind of help would you have liked to have had?
3. Have you ever been asked to help someone in their searches?
If “yes”:
 - (a) How often? [note, if the participant works as a reference librarian, ask them to select two recent cases without discussing any identities]
 - (b) Can you describe one such incident? This should include motivation, methods, and tools used.
4. If at least one of the answers for questions 2 and 3 was a “yes”:
 - (a) In searching for information for somebody else, could you imagine a better set of tools than what you used? If “yes”, what would be on your wish list?
5. Do you ever do co-searching? That is, you and one or more others search separately and then bring the results together for everyone’s benefit?
If “yes”:
 - (a) Describe one such incident. This should include motivation, methods, and tools used.
6. Now let’s talk about working in groups in general. How often do you work on group projects?
7. Give some examples of group projects that you were involved in.

8. [For a selected example, ask the following questions:]
- (a) Why did you decide to work together?
 - (b) How many people were involved? Did the size of the group change during the process? If “yes”, why?
 - (c) What were their roles? Was there a leader of the group?
 - (d) How long did the collaboration last?
 - (e) What tools or methods did you use to collaborate? Email, chat, wiki, blogs, meetings, phone, other?
 - (f) Was there a part in your project that included searching for information from external sources?
If “yes”:
 - i. Who did these searches and how did they fit in the overall project?
 - ii. If more than one person did searching, then describe (1) why, (2) how, and (3) what tools they used.
 - (g) Do you think you could have done all the work by yourself in the collective time that the group took? Why or why not?
 - (h) If you had to work on the same project again, what would you do differently and why? This may include picking a different team.

Appendix C

Laboratory Study Recruiting

Coagmento

Research Study

Study Signup Page 1 of 2

Please read these instructions carefully.

- This form allows you to *request your participation* in the study. All the fields are required.
- The study will be conducted at SILS (Manning, next to Lenoir) on the UNC Campus at the time of your choosing.
- This study requires a *team of two*. You need to sign up in pairs.
- Both of you should have worked on some project before (e.g., a class assignment).
- You need to sign up for *two sessions*, which are one to two weeks apart. Available time slots are given at the bottom of this page for reference.
- The study will involve using an experimental system, called *Coagmento* (a browser plug-in) while surfing the Web.
- You will be asked to fill in some questionnaires and interviewed briefly during this study.
- No identifying information about you will be shared.
- You will be paid total **\$25 (per person) for two sessions** (thus, a team will receive **\$50**).
- The best performing team will win **iPod Shuffles** (two per team).

Participant-1 details	Participant-2 details
First name <input type="text" value="Bugs"/>	First name <input type="text" value="Daffy"/>
Last name <input type="text" value="Bunny"/>	Last name <input type="text" value="Duck"/>
Email <input type="text" value="bugs@unc.edu"/>	Email <input type="text" value="daffy@unc.edu"/>

Briefly, describe a project that you two have worked together in the past, or are working on now.

We used to hunt for carrots together! :)|

Next->

Available time slots for this study are given below. You will be able to pick your slots on the next page.

If none of these slots work for you exactly as they are, fill in your details, leave us comments on the next page, and we may be able to find something that could work for you!

Week of September 14th - Click here to expand or collapse the time slots

Week of September 21st - Click here to expand or collapse the time slots

>> Questions? [Contact us](#) <<

Figure C.1: Study recruiting page 1 of 2.

Coagmento

Research Study

Study Signup Page 2 of 2

Choose the **earliest possible time-slots** from the following. Remember, you both need to show up together for both the slots, and the **slots need to be one to two weeks apart**. If none of the slots work for you exactly as they are, we may be able to find you one with the time that is more suitable to you. In that case, please select the closest slot(s) and leave us a note in the comments box.

Step-1: Select a slot for Session-1

Note that some of the selections below may not result in any valid choices on the right.

Week of September 14th - Click here to expand or collapse the time slots

- ☐ Monday, 2009-09-14, 11am-12:30pm
- ☐ Monday, 2009-09-14, 2pm-3:30pm
- ☐ Tuesday, 2009-09-15, 9am-10:30am
- ☐ Tuesday, 2009-09-15, 11am-12:30pm
- ☐ Tuesday, 2009-09-15, 6pm-7:30pm
- ☐ Wednesday, 2009-09-16, 9am-10:30am
- ☐ Wednesday, 2009-09-16, 11am-12:30pm
- ☐ Wednesday, 2009-09-16, 2pm-3:30pm
- ☐ Thursday, 2009-09-17, 9am-10:30am
- ☐ Friday, 2009-09-18, 11am-12:30pm
- ☐ Friday, 2009-09-18, 2pm-3:30pm
- ☐ Saturday, 2009-09-19, 9am-10:30am
- ☒ Sunday, 2009-09-20, 9am-10:30am

Week of September 21st - Click here to expand or collapse the time slots

Step-2: Possible slots for Session-2

- ☐ Monday, 2009-09-28, 9am-10:30am
- ☐ Sunday, 2009-09-27, 9am-10:30am
- ☐ Sunday, 2009-09-27, 11am-12:30pm
- ☐ Sunday, 2009-09-27, 2pm-3:30pm
- ☐ Sunday, 2009-09-27, 4pm-5:30pm
- ☐ Sunday, 2009-09-27, 6pm-7:30pm
- ☐ Monday, 2009-09-28, 11am-12:30pm
- ☐ Monday, 2009-09-28, 2pm-3:30pm
- ☐ Monday, 2009-09-28, 4pm-5:30pm
- ☐ Monday, 2009-09-28, 6pm-7:30pm
- ☐ Tuesday, 2009-09-29, 9am-10:30am
- ☐ Tuesday, 2009-09-29, 11am-12:30pm
- ☐ Tuesday, 2009-09-29, 2pm-3:30pm
- ☐ Tuesday, 2009-09-29, 4pm-5:30pm
- ☐ Tuesday, 2009-09-29, 6pm-7:30pm
- ☐ Wednesday, 2009-09-30, 9am-10:30am
- ☐ Wednesday, 2009-09-30, 11am-12:30pm
- ☐ Wednesday, 2009-09-30, 2pm-3:30pm
- ☐ Wednesday, 2009-09-30, 4pm-5:30pm
- ☐ Wednesday, 2009-09-30, 6pm-7:30pm

Figure C.2: Study recruiting page 2 of 2.

Appendix D

Laboratory Study Demographic Questionnaire

1. Your age: ---
2. Gender: __ Female __ Male
3. Your program of study: __ Undergraduate __ Graduate __ Other
4. What is your major course of study or profession? -----
5. Which operating system do you use most frequently?
__ Mac __ Windows __ Linux __ Other
6. Which browser do you use most frequently?
__ Firefox __ Internet Explorer __ Safari __ Other
7. How would you describe your search experience?
(Very Inexperienced) 1 2 3 4 5 6 7 (Very Experienced)
8. How often do you search the Web?
__ Occasionally
__ 1-3 searches per day
__ 4-6 searches per day
__ 7-10 searches per day
__ More than 10 searches per day
9. How often do you work on a project with others?
__ Daily __ Weekly __ Monthly __ Less than monthly
10. How many collaborative projects did you work on during the past year? ---
11. How much did you enjoy working on these collaborative projects?
(Not at all) 1 2 3 4 5 6 7 (Very much)
12. How successful were these collaborative projects?
(Not at all) 1 2 3 4 5 6 7 (Very much)

13. How often do you use text messaging?
- Occasionally
 - 1-3 messages per day
 - 4-6 messages per day
 - 7-10 messages per day
 - More than 10 messages per day

Appendix E

Query Re-usage Analysis

Four particular questions about query re-usage (within individuals) and overlap (between individuals) seemed interesting and appropriate to investigate from the data collected through this study. These explorative questions (EQs) and the corresponding observations and analyses follow.

EQ1. How often do people re-use their own queries?

Figure E.1 plots the queries used for each task and session, along with what proportion of them were unique and overlapping. As we can see, a large portion of queries for a given session was already used in that session (about 20 minutes in length). This is also reflected in Table E.1, where the query re-use statistics are reported. Both the tasks had on average about 60% of query re-use.

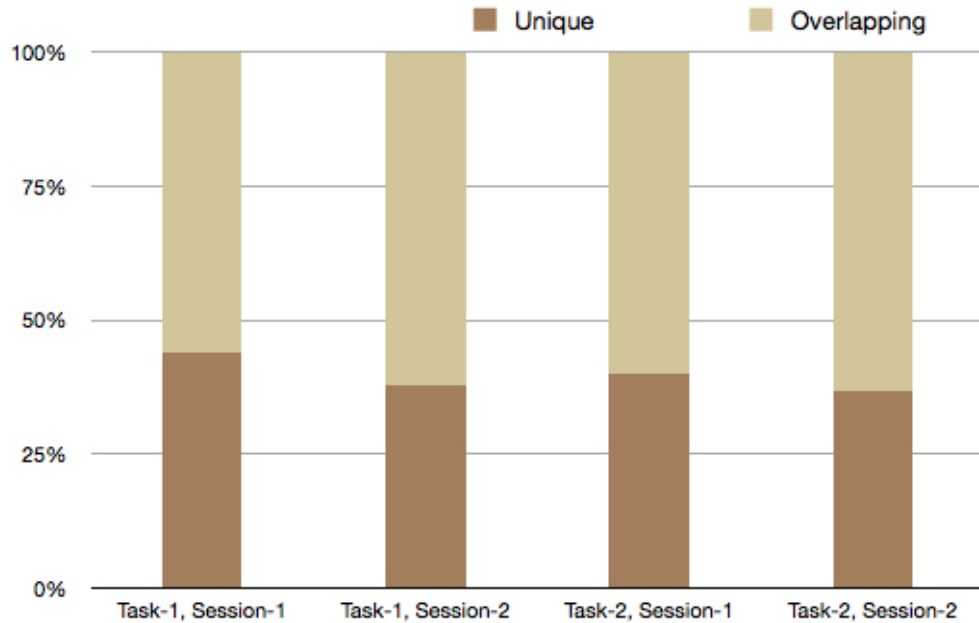


Figure E.1: Query re-usage in a given session for a given task.

It is also interesting to note that, for both the tasks, during the second session, there was more re-usage of queries than during the first session.

EQ2. How often do people re-use their queries from the previous session?

One of the interesting subgoals in this study was to look at browsing and query re-use

Table E.1: Task and session-wise individual query re-usage statistics.

	Session-1	Session-2	Average
Task-1	54.23%	60.41%	57.32%
Task-2	59.28%	61.07%	60.18%
Average	56.76%	60.74%	

across multiple sessions. It was found that only about 5-10% of the queries used in the second session were repeats from the first session. However, when the matching criteria were expanded to include subqueries (e.g., “economics recession” is a subquery of “economics recession US”), a much larger re-usage portion was found. In fact, for Task-2, nearly half of the queries were found to be repeats (i.e., as the same exact query or a subquery) from the first session. This is shown in Figure E.2.

Further analysis showed that, for Task-2, more than half of the queries had “social networking” in them. This may be due to the fact that almost all the facets in this task also had “social networking” as a sub-facet; the participants found it difficult to investigate those facets without the context of social networking. This also became apparent in the interviews conducted after the tasks. For Task-1, on the other hand, the participants could run fairly independent queries for covering different facets, such as “unemployment stats” and “recession causes”.

EQ3. What proportion of queries overlap across different people for the same task?

Figure E.3 shows the portion of queries used by each participant that were also used by some other participant for the same task. Similar to above, subquery matches were looked at in addition to exact query match. We can see that for Task-2, the participants had a much better agreement on what queries were reused. This confirms the justification given at the end of the previous question analysis.

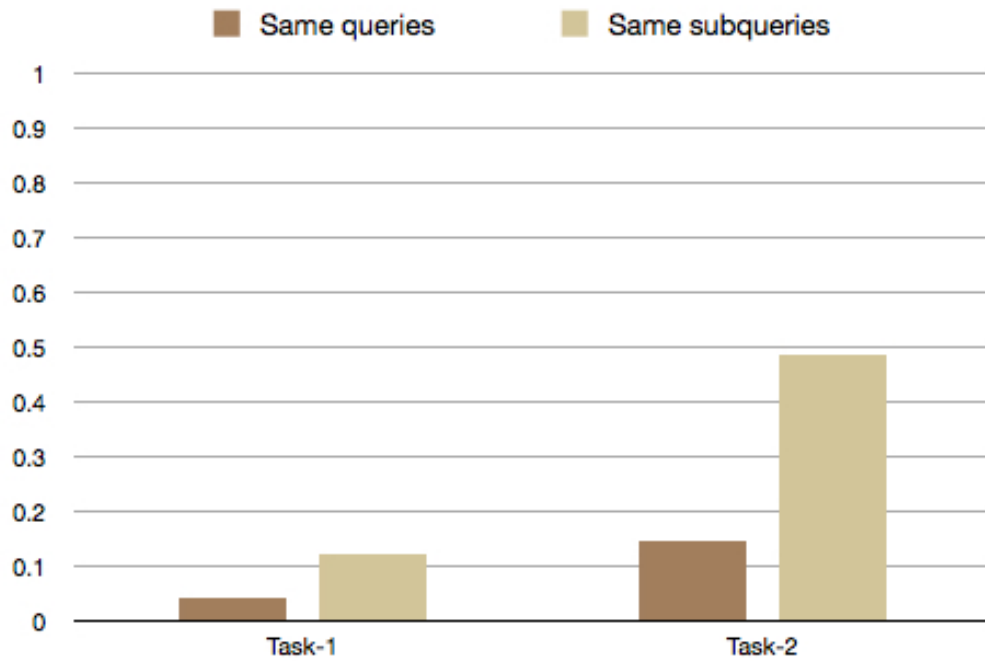


Figure E.2: Second session query and subquery re-usage rates from the first session.

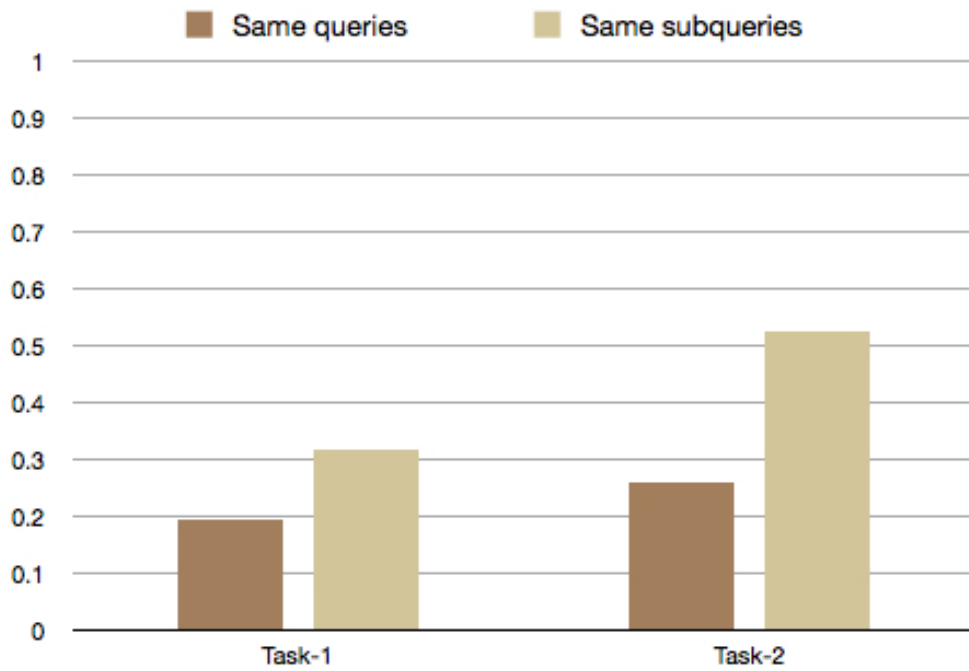


Figure E.3: Average query and subquery re-usage proportion for a participant with respect to other participants.

EQ4. What proportion of queries is similar across different people of different teams for the same task?

Instead of simply looking at exact matching queries, the closeness between two given queries was also looked at. To find this closeness, the Edit Distance¹ measure was used. The results are plotted in Figure E.4 and E.5. In these figures, the X-axis shows the edit distance between two queries, and Y-axis shows the number of queries. Thus, for Task-1, there were 528 queries that had a closest query with distance zero (exact match), 162 queries that had a closest query with distance one, and so on. For simplicity, edit distance only up to 20 is shown in both the graphs.

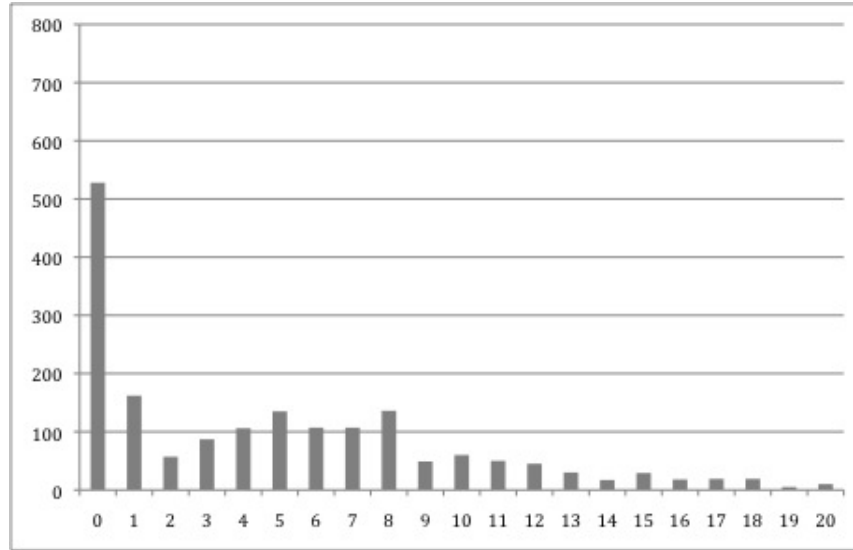


Figure E.4: Edit Distance among the queries for Task-1. X-axis shows Edit Distance between a pair of queries, and Y-axis shows number of queries.

Once again, we find that many queries that our participants used for a given task were the same or very similar to the queries other(s) have used.

It was also found that in the case of Task-2, there was a greater agreement among the participants in formulating the queries, as compared to Task-1 (e.g., 725 queries with zero edit distance for Task-2 vs. 528 for Task-1).

From an analysis of query usage of these 84 participants working on exploratory tasks over two sessions, support was found for query re-usage for individuals, and high overlap among the queries of multiple participants for a given task. Such observations and analyses about query usage and re-usage confirm that people with the same information need tend to express their information need in the same/similar way. This is a driving motivation for collaborative filtering work and query assistance/suggestion.

¹The edit distance between two strings the number of operations (addition, deletion, changing of characters) required to transform one of them into the other.

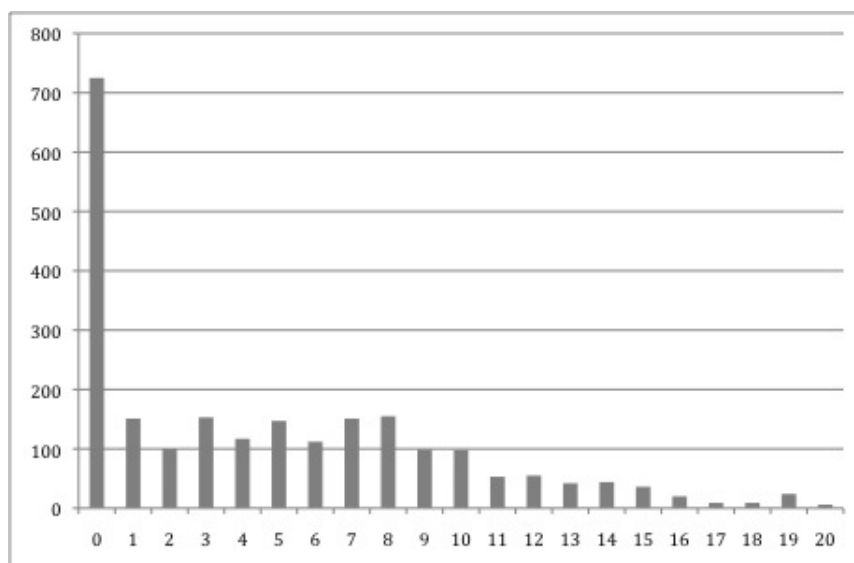


Figure E.5: Edit Distance among the queries for Task-2. X-axis shows Edit Distance between a pair of queries, and Y-axis shows number of queries.

The substantial reuse and overlap demonstrate that such techniques may be even more useful for exploratory searching.

Appendix F

Interaction Effects

Dependent Variable: Index					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	26.523	11	2.411	2.806	.002
Intercept	7362.737	1	7362.737	8567.016	.000
Condition	2.998	2	1.499	1.744	.176
Session	10.410	1	10.410	12.113	.001
Task	7.188	1	7.188	8.363	.004
Condition * Session	4.017 ^a	2	2.009	2.337	.098
Condition * Task	.374	2	.187	.218	.805
Session * Task	1.259	1	1.259	1.465	.227
Condition * Session * Task	.276	2	.138	.160	.852
Error	278.455	324	.859		
Total	7667.714	336			
Corrected Total	304.977	335			

a. R Squared = .087 (Adjusted R Squared = .056)

Figure F.1: Interaction effects between tasks and sessions for post-task reported awareness questionnaire (measured by an index). No significance difference between conditions.

Dependent Variable: Index					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11.923	11	1.084	1.400	.171
Intercept	3591.376	1	3591.376	4637.804	.000
Condition	3.393	2	1.697	2.191	.113
Session	3.407	1	3.407	4.399	.037
Task	3.274	1	3.274	4.228	.041
Condition * Session	1.287 ^a	2	.643	.831	.437
Condition * Task	.218	2	.109	.141	.869
Session * Task	.153	1	.153	.197	.657
Condition * Session * Task	.191	2	.095	.123	.884
Error	250.896	324	.774		
Total	3854.194	336			
Corrected Total	262.819	335			

a. R Squared = .045 (Adjusted R Squared = .013)

Figure F.2: Interaction effects between tasks and sessions for post-task reported cognitive load (measured by an index). No significance difference between conditions.

Dependent Variable: Index					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	29.720	5	5.944	2.163	.061
Intercept	4986.834	1	4986.834	1814.411	.000
Condition	27.473	2	13.737	4.998	.008
Session	.331	1	.331	.121	.729
Condition * Session	1.915	2	.958	.348	.706
Error	445.250	162	2.748		
Total	5461.805	168			
Corrected Total	474.970	167			

a. R Squared = .063 (Adjusted R Squared = .034)

Figure F.3: Interaction effects between sessions for end-session ease of use and satisfaction questionnaire (measured by an index). No significance difference between sessions.

Dependent Variable: Index					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	6.128	5	1.226	1.824	.111
Intercept	5234.375	1	5234.375	7790.938	.000
Condition	5.751	2	2.876	4.280	.015
Session	.271	1	.271	.404	.526
Condition * Session	.105	2	.053	.078	.925
Error	108.840	162	.672		
Total	5349.344	168			
Corrected Total	114.968	167			

a. R Squared = .053 (Adjusted R Squared = .024)

Figure F.4: Interaction effects between sessions for end-session engagement questionnaire (measured by an index). No significance difference between session.