The Information Retrieval Series Volume 34

Series Editor

W. Bruce Croft

Editorial Board

ChengXiang Zhai
Maarten de Rijke
Nicholas J. Belkin
Charles Clarke

Chirag Shah

Collaborative Information Seeking

The Art and Science of Making the Whole Greater than the Sum of All

Springer

Chirag Shah
School of Communication & Information
The State University of New Jersey
New Brunswick, NJ, USA

ISSN 1387-5264 The Information Retrieval Series
ISBN 978-3-642-28812-8 ISBN 978-3-642-28813-5 (eBook)
DOI 10.1007/978-3-642-28813-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012942855

ACM Computing Classification (1998): H.3, K.4

© Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my late father and eternal role model—
Rajendrakumar Shah

Foreword

Information seeking is a term that describes a collection of fundamental human processes that enable survival and growth. Some elements of information seeking are involuntary and continuous elements of consciousness, but most studies of information seeking focus on conscious processes that tend to be grouped under the name 'search'. In the 20th century, industrialization led to an informated society, which in turn has made search for information one of the basic skills of 21st century modernity. Information work is seldom solitary and most information is selectively shared at different stages of the information life cycle. Not only do people collaborate to create or analyze information, but they often collaborate on the seeking processes. This book focuses on theory, systems, and evaluation for the evolving study of collaborative information seeking.

Research and development in the information, library, and computer science domains led to great advances in search services. Today, the easy cases of search are well-served by search engines, e.g., known item searches, single facet searches, and searches for pointers to content. If the answer to a question can be expressed by a single info object (e.g., a page, a figure) then search engines can find that object. What remains problematic for these cases are assessing the veracity of the source and the appropriate interpretation and the use of information objects. However, there is much research to be done on search for open-ended, exploratory problems where answers require many information objects and points of view. The answer to a thesis question, which medical treatment strategy to select, how to invest for your retirement, which university to attend, the effects of social media on adolescent behavior, or similar questions, require people to invest more time, marshal more systems and services, and orchestrate the interaction of multiple subprocesses. Increasingly, the actions of other people add new dimensions to the multiple steps associated with multi-session, collaborative information seeking. This kind of information seeking requires a lot more support than current search engines provide.

The subtleties of human–information interaction done alone are complex: mapping needs to words (query tokens or displayed menu metadata), managing information tools and services, scanning candidate result snippets to evaluate whether to invest further examination, consuming documents to assess relevance and construct

meaning, synthesizing these assessments over multiple consumptions to reach some conclusion or decision, and monitoring the process over perhaps long time intervals. Doing these activities with other people introduces possibilities for improved effectiveness but adds considerable collaborative load costs to share, achieve mutual understanding, form consensus, and monitor or document the collaborative information seeking process. We are only beginning to understand the tradeoffs that accrue between effectiveness and efficiency as people increasingly are able to seek information together. In this volume, Chirag Shah provides both a framework for thinking about and studying collaborative information seeking, and a roadmap for building systems that improve both effectiveness and efficiency.

Chapel Hill, NC, USA Gary Marchionini

Preface

We live in a society where information is, without a doubt, a powerful force. This statement may sound like a cliche, but it always amazes me how often we forget. May be that's the purpose (or should be) of the technology that surrounds such information. The other aspect of information technology that amazes me is the fact that it is so new, considering human history. The clock that tells us the precise time of the day dates back only to the sixteenth century. The base 10 numbering system is only 500 years old, and mechanical devices used to calculate and present information have existed for only a couple of centuries. Of course, today when we say "information technology", we are probably thinking about computers and other digital devices, and they are merely a few decades old.

What intrigues me the most is how we have been able to integrate such new concepts and technologies with long-standing human behavior. Take for example, working in groups and living as a society. This behavior has proven to be extremely important for the survival and prosperity of our species. Back in the days of hunting together to today's office work, mankind has understood the need to work and thrive together. It is this behavior—the one of collaborating with each other—that has made it possible to achieve great feats in the history. How else can one man (or woman)[1] build the pyramids[2] or crack the human genetics code.

Of course, not all problems call for people working together. While Einstein had help and drew inspirations from others, he did come up with many significant findings himself. Leonardo de Vinci and Picasso, similarly, worked alone. Claude Shannon, considered to be the father of digital information age, was known to have worked in solitude behind closed doors. But let's put geniuses aside and talk about the remaining 99.99 % of us (which of course, still includes a lot of smart people). We do, often need to work in collaboration. I'm sure even Einstein needed help placing his furniture in his Princeton house; he was a genius, not a superhuman!

[1] I'm not going to worry about being politically correct in this book and assume that the savvy reader understands he or she being equal.

[2] Though the ancient Egyptian workers were forced to work on pyramid construction on such "forced collaboration". More on this later.

This book is about those times when people work in collaboration—an eternal human behavior, in the light of new and innovative technologies in the information age that we live in. More precisely, it is concerned about situations pertaining to information retrieval/seeking/sense-making where people are collaborating or should be collaborating.

One may ask—Why this book? Why now? There is a simple two-fold answer to both these questions. Using technology to understand and support collaborative behavior has been around for a while—what is known as Computer-Supported Cooperative Work (CSCW), but it is in the recent years that we have seen more specialized attention given to applying CSCW methods and frameworks for information seeking situations. On the other hand, the field of Information Retrieval (IR, or broadly speaking IS, information seeking) has found (or re-realized) the importance of considering social and collaborative aspects of search, synthesis, and information use.

This has led to a newly developed interest in the field that is still emerging at the intersection of several other well-established fields, including CSCW, IR, HCI, and social media/networking. This book as an attempt to introduce the relatively young domain of collaborative information seeking (CIS) research by discussing how it came to be, what it currently offers, and where it is headed next. The best part is that we all get to define and contribute to this future.

Personally, my journey on this path started during the summer of 2007 when I was an intern at FXPAL, working with Gene Golovchinsky and Jeremy Pickens. Back then, we worked on something called Collaborative Exploratory Search (CES), and argued that IR systems need to have "smart" components that could mediate collaborative activities and produce results that are "better" than any individual IR process, and we succeeded with at least one kind of situation (time-limited, recall-oriented task with two people collaborating under assumed roles). We did continue this work further by identifying more situations and defining other roles, but as I returned to UNC and resumed working on my dissertation, I started moving in the direction of user-mediated collaboration. My dissertation provided a framework (among other things) for studying and supporting user-focused CIS. I have carried on working on various aspects of CIS (both user and system sides) as a faculty at Rutgers University. In the meantime, I have also participated in a number of professional events around CIS, including half a dozen workshops—two of which I co-organized.

This book is a culmination of all of these experiences, and while they have made me biased on the topic, I have tried my best to include others' views as well. In the end, my hope is that those working in this domain, and the larger field of IR see this book as a record of modern day CIS research that has tried to incorporate many view-points and contributions to inform those looking for a comprehensive treatment of this topic, along with the wonderful opportunities (and challenges) it presents.

New Brunswick, NJ, USA Chirag Shah

Acknowledgements

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

- Gene Golovchinsky and Jeremy Pickens at FXPAL, California, for introducing me to the area of collaborative exploratory search during my internship in summer 2007.
- The National Science Foundation, for funding my dissertation studies and research with IIS grant # 0812363.
- Yahoo! for providing funding for the development of Coagmento through Campus Innovation Award (2011–2012).
- Rutgers University PhD students Roberto Gonzalez-Ibanez and Muge Haseki, who have not only been working with me on various issues of collaborative information seeking (CIS), but also contributed to parts of this book. Specifically, Roberto wrote a section on affective dimension of CIS, and Muge did Appendix D on computer-mediated communication (CMC).
- My PhD 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—Diane Kelly, Barbara Wildemuth, and late Deborah Barreau from UNC Chapel Hill, and Susan Dumais from Microsoft Research.
- My colleagues at Rutgers University, especially Nick Belkin, Paul Kantor, Michael Lesk, Marie Radford, and Mark Aakhus who enriched my understanding of CIS and the fields of information and communication through many conversations and mentoring sessions.
- A number of external advisors who have either commented on different drafts of my work (articles, proposals, this manuscript) or had very helpful discussions with me on this topic—Madhu Reddy from Penn State; Michael Twidale from UIUC; Rob Capra from UNC Chapel Hill; Jonathan Foster from University of Sheffield; Preben Hansen from Swedish Institute of Computer Science; Susan

Dumais, Merrie Morris, Ryen White, Paul Bennett, and Jaime Teevan from Microsoft Research; and Dan Russell and Ed Chi from Google.

- My late father Rajendrakumar, mother Sneha, and sister Shweta, 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 book. Even in the darkest hour of toil and frustration, she had faith in me and my efforts. Lori also provided valuable feedback on early drafts of this book, as well as beta versions of Coagmento. Without her, this work could not have been possible.

Contents

Part I Introduction

1 Introduction . . . 3
1.1 Collaboration in Information Seeking Situations . . . 3
1.2 Information Retrieval, Seeking, and Collaboration . . . 4
1.2.1 Collaboration to Help Information Seeking . . . 4
1.2.2 Information Seeking to Help Collaboration . . . 5
1.3 CIS Scenarios . . . 6
1.3.1 What/How We Already Do with Regard to CIS . . . 6
1.3.2 What/How We Would Like to do During CIS . . . 7
1.4 Structure of the Book . . . 8
References . . . 9

2 Collaboration . . . 11
2.1 Definitions and Models of Collaboration . . . 11
2.2 Principles of Collaboration . . . 16
2.3 Process of Collaboration . . . 17
2.4 Limitations of Collaboration . . . 18
2.5 Collaboration in the Context of Information-Intensive Tasks . . . 20
2.6 How the Concept of Collaboration Relates to CIS . . . 21
References . . . 23

3 Collaborative Information Seeking (CIS) in Context . . . 25
3.1 Introduction . . . 25
3.2 Collaborative Information Retrieval (CIR) and Co-search . . . 26
3.3 Co-browsing or Collaborative Navigation . . . 28
3.4 Social Search . . . 30
3.5 Collaborative Filtering . . . 32
References . . . 36

Part II Conceptual Understanding of CIS

4 Frameworks for CIS Research and Development 41
4.1 Introduction 41
4.2 Space and Time Aspects of CIS 41
4.3 User–Source–Time Configuration for CIS 42
4.4 Control, Communication, and Awareness in a CIS Environment . . 44
4.4.1 Control 45
4.4.2 Communication 46
4.4.3 Awareness 46
4.4.4 Importance of Control, Communication, and Awareness in CIS Systems 49
4.5 Nature and Level of Mediation 50
4.5.1 System/Algorithmically Mediated Collaboration 50
4.5.2 User/Interface Mediated Collaboration 51
4.6 Dimensions of Collaborative Systems 51
4.7 Summary 56
References 57

5 Toward a Model for CIS 61
5.1 Models for Collaboration 61
5.2 Models for Information Seeking 63
5.3 Models for User and System Interaction 68
5.3.1 Degree of Involvement 72
5.3.2 Explicitness of Collaboration 73
5.4 Toward a Cognitive Model for CIS 74
5.4.1 Capturing Affective Dimension of CIS 74
5.4.2 Extending Information Seeking Processes (ISP) Model for Collaboration 77
5.5 Summary 82
References 83

Part III CIS Systems, Applications, and Implications

6 Systems and Tools for CIS 89
6.1 Costs, Challenges, and Design Implications for CIS Systems . . . 89
6.2 System-Mediated Collaboration 91
6.2.1 Cerchiamo 91
6.2.2 Querium 92
6.3 User-Mediated Collaboration 93
6.3.1 Ariadne 93
6.3.2 SearchTogether 95
6.4 A Case Study in Designing a CIS Solution with Coagmento 96
6.4.1 Design 96
6.4.2 Development 104
6.4.3 Deployment 109

6.5 Future Development . . . 112
References . . . 113

7 Evaluation . . . 115
7.1 Challenges for Evaluating CIS Systems and Situations . . . 115
7.2 Evaluation Methodologies . . . 116
7.2.1 Evaluating Usability with User Studies . . . 116
7.2.2 System-Based Training-Testing . . . 117
7.2.3 Qualitative Evaluation . . . 117
7.2.4 Task or Application Based Evaluation . . . 117
7.3 System Measures . . . 118
7.3.1 Precision, Recall, and F-measure . . . 118
7.3.2 Coverage . . . 119
7.3.3 Effectiveness . . . 120
7.3.4 Likelihood of Discovery . . . 120
7.3.5 Diversity . . . 121
7.4 User Measures . . . 121
7.4.1 Collaborative Aptitude . . . 121
7.4.2 Usability . . . 123
7.4.3 Cognitive Load . . . 124
7.4.4 Engagement . . . 124
7.4.5 Awareness . . . 125
7.4.6 Affects/Emotions . . . 125
7.5 Case Study . . . 126
7.5.1 Subjects . . . 127
7.5.2 System . . . 127
7.5.3 Session Workflow . . . 128
7.5.4 Conditions . . . 128
7.5.5 Task . . . 129
7.5.6 Results and Discussion . . . 130
7.5.7 Conclusion . . . 134
7.6 Summary . . . 135
References . . . 135

8 Conclusion . . . 139
8.1 CIS—Looking Back, Looking Forward . . . 139
8.2 Theoretical Issues . . . 139
8.3 Experimental and Practical Issues . . . 141
8.4 Summary . . . 143
References . . . 143

Appendix A Ten Stories of Five *C*s . . . 145
A.1 Introduction . . . 145
A.2 A Model of Collaboration . . . 145
A.3 Scenarios . . . 147
A.3.1 Paying Taxes . . . 147
A.3.2 What's for Dinner? . . . 148

A.3.3 School Elections . . . 149
A.3.4 Group Project . . . 149
A.3.5 VidArch . . . 150
A.3.6 Searching in a Library . . . 151
A.3.7 Planning a Trip . . . 152
A.3.8 Discovery of SARS Virus . . . 153
A.3.9 Factory Line . . . 154
A.3.10 May Day Parade . . . 155
A.4 Conclusion . . . 156
References . . . 157

Appendix B Brief Overview of Computer-Supported Cooperative Work (CSCW) . . . 159
B.1 Organizing CSCW Systems and Environments . . . 159
B.2 Control, Communication, and Awareness in CSCW . . . 160
B.2.1 Control . . . 160
B.2.2 Communication . . . 162
B.2.3 Awareness . . . 162
B.3 Groupware Systems . . . 164
B.4 Summary . . . 165
References . . . 166

Appendix C Brief Overview of Computer-Supported Collaborative Learning (CSCL) . . . 169
C.1 Theories . . . 169
C.2 Practice and Applications . . . 170
C.3 Summary . . . 171
References . . . 171

Appendix D Brief Overview of Computer-Mediated Communication (CMC) . . . 173
D.1 Social Presence Theory . . . 173
D.2 Media Richness Theory . . . 174
D.3 Social Information Processing (SIP) Theory . . . 175
D.4 Social Identification/De-individuation (SIDE) Theory . . . 176
D.5 Hyperpersonal Communication Model . . . 176
D.6 Summary . . . 177
References . . . 177

Glossary . . . 179

Index . . . 183

List of Abbreviations

CES	Collaborative Exploratory Search
CIB	Collaborative Information Behavior
CIR	Collaborative Information Retrieval
CIS	Collaborative Information Seeking
CMC	Computer-Mediated Communication
CS	Computer Science
CSCL	Computer-Supported Cooperative Learning
CSCW	Computer-Supported Cooperative Work
HCI	Human–Computer Interaction
IR	Information Retrieval
IS	Information Seeking
LIS	Library & Information Science
PIM	Personal Information Management

List of Figures

Fig. 1.1 Contextual depiction of Collaborative Information Seeking (CIS) . 5
Fig. 2.1 A set-based model of collaboration. An inner set is essential to or supports the outer set 13
Fig. 2.2 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 15
Fig. 3.1 Depiction of CIS and related topics using the dimensions of human–system and explicit–implicit collaboration 26
Fig. 3.2 A typical scenario of information seeking in an IR environment . . 33
Fig. 3.3 Content-based information filtering 33
Fig. 3.4 Collaborative information filtering 34
Fig. 4.1 Space–time dimensions for library activities [15]. Permit pending . 42
Fig. 4.2 Various collaborative activities and systems along space and time dimensions 43
Fig. 4.3 A framework for CIS environments (part 1 of 2) 52
Fig. 4.4 A framework for CIS environments (part 2 of 2) 53
Fig. 5.1 Steps to collaboration (http://www.empowerment.state.ia.us/files/annual_reports/2001/Collaboration.pdf) 62
Fig. 5.2 A set-based model of collaboration. An inner set is essential to or supports the outer set 62
Fig. 5.3 Four layer model of information seeking centered around information access and organization 65
Fig. 5.4 Extension of the four layer model of information access and organization 66
Fig. 5.5 A model for Collaborative Information Seeking (CIS) extended from the model in Fig. 5.3 68
Fig. 5.6 Saracevic's model of stratified interaction 70
Fig. 5.7 Looking at collaboration with the amount of user and system involvement 72

Fig. 5.8 Looking at collaboration with the amount of explicitness or intention from user and system 73
Fig. 5.9 Collaborative ISP stages for session-1 79
Fig. 5.10 Positive/negative feelings, and affective relevance for session-1 . . 80
Fig. 5.11 Collaborative ISP stages for session-2 80
Fig. 5.12 Positive/negative feelings, and affective relevance for session-2 . . 81
Fig. 6.1 Cerchiamo's architecture. Printed with permission 92
Fig. 6.2 Querium interface: (*1*) top area for new queries, (*2*) navigation sidebar, (*3*) main results, and (*4*) document view. Printed with permission 93
Fig. 6.3 A search visualization in Ariadne. Printed with permission 94
Fig. 6.4 Screenshot of SearchTogether. (*a*) Integrated messaging, (*b*) query awareness, (*c*) current results, (*d*) recommendation queue, (*e*)–(*g*) search buttons, (*h*) page-specific metadata, (*i*) toolbar, (*j*) browser. Printed with permission 95
Fig. 6.5 Coagmento v1: design version 98
Fig. 6.6 Toolbar provided on top of the document being viewed with Coagmento 99
Fig. 6.7 Coagmento v2: development version 105
Fig. 6.8 Snippets collection window 105
Fig. 6.9 Annotations window 106
Fig. 6.10 Participants during a study session 107
Fig. 6.11 Coagmento v3: deployment 109
Fig. 7.1 A framework for IR-focused system-based evaluation measures . . 119
Fig. 7.2 A version of Coagmento used for the study 127
Fig. 7.3 Study session workflow 128
Fig. 7.4 Experimental setup for four different conditions 129
Fig. 7.5 Depiction of coverage by various conditions (drawn to scale) 132
Fig. 7.6 Query distance for $C2_{\text{ARTIFICIAL}}$, $C4_{\text{SAME_ROOM}}$, and $C5_{\text{REMOTE}}$ using Lavenshtein algorithm 133
Fig. 8.1 CIS as an interdisciplinary field, along with some of the issues stemming from different parts of it 140
Fig. A.1 A model for collaboration 146
Fig. B.1 CSCW/groupware matrix showing how different collaborative activities and systems fit across time and space dimensions. Courtesy of Momo54 (http://en.wikipedia.org/wiki/User:Momo54) 160

List of Tables

Table 2.1 Various group activities and examples 20
Table 4.1 CIS in the context of user–source–time dimensions 43
Table 4.2 Differentiating between intent and activeness 54
Table 5.1 Information seeking strategies (Belkin et al., 1995) 71
Table 5.2 Summary of participants' feelings and affective relevance through ISP stages 82
Table 6.1 Summary of responses by 84 participants on ease of use and satisfaction questionnaire for the lab study 108
Table 6.2 Summary of responses by 19 participants on ease of use and satisfaction questionnaire for the field study 110
Table 6.3 Summary of responses by 8 participants on ease of use and satisfaction questionnaire for the participatory design phase II study 111
Table 7.1 Experimental conditions 129
Table 7.2 Summary of various universes used in our analysis 130