**Thilo Böhm**

# System-Mediated Support of Explicit Collaboration in Information Retrieval

**Mathematik und Informatik**

Dissertation



FernUniversität in Hagen

# System-Mediated Support of Explicit Collaboration in Information Retrieval

dem Fachbereich Mathematik und Informatik
der FernUniversität in Hagen
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
vorgelegte

DISSERTATION

von

Thilo Böhm, M.Sc.

aus Berlin

Hagen, März 2016

| | |
|---|---|
| Referent: | Prof. Dr.-Ing. Matthias Hemmje |
| Korreferent: | Prof. Dr.-Ing. Norbert Fuhr |
| Tag der Einreichung: | 01.03.2016 |
| Tag der mündlichen Prüfung: | 10.06.2016 |

## Vorwort

Die vorliegende Arbeit entstand im Rahmen einer externen Promotion am Lehrgebiet Multimedia- und Internetanwendungen der FernUniversität in Hagen, welches halbjährig Dokorandenseminare in Hagen, Darmstadt oder Köln organisierte und so den Doktoranden die Möglichkeit gab, regelmäßig über den Stand ihrer Arbeiten zu berichten, diese zu diskutieren und sich auszutauschen. Hierdurch wurde eine wertvolle Möglichkeit geboten, sich nebenberuflich in aktuellen Forschungsthemen zu engagieren. Für den Autor dieser Dissertation ergab sich dadurch die Chance, Fragegestellungen, die einem im beruflichen Alltag begegneten, mit wissenschaftlichen Methoden zu formalisieren und in einer sonst nicht möglichen Tiefe zu untersuchen.

Ich bedanke ich mich bei Prof. Dr.-Ing. Matthias Hemmje für seine Betreuung und die Möglichkeit zur Promotion an der FernUniversität in Hagen. Die fachkundige Betreuung, wertvollen Hinweise und vorausschauende Planung des Promotionsvorhabens ermöglichten es, diese Arbeit als externer Doktorand anzufertigen. Mein Dank gilt weiterhin Dr. Claus-Peter Klas für die vielen motivierenden Gespräche und fachlichen Diskussionen sowie wertvollen Anregungen. Weiterhin möchte ich den Studenten Robert Dronsgalla, Christian Steiner sowie Jan Lagerpusch danken, die am Lehrgebiet Multimedia- und Internetanwendungen ihre Abschlussarbeiten geschrieben haben und die ich bei der Erstellung ihrer Abschlussarbeiten betreuen durfte. Ihre Beiträge haben geholfen, die Evaluationssoftware zu erstellen, die für diese Arbeit eingesetzt wurde.

# Kurzfassung

Traditionell betrachtet die Forschung im Bereich des Information Retrieval (IR) die Informationssuche als überwiegend individuelle Aktivität. Die gängige Annahme ist, dass eine einzelne Person Suchanfragen stellt und die Suchergebnisse beurteilt. Jedoch zeigen neuere Studien, dass kollaboratives IR ein häufiger Bestandteil im professionellen Arbeitsalltag ist, in dem die Teamarbeit betont wird. Teammitglieder arbeiten dabei explizit zusammen und jedes Teammitglied steuert individuelle Kenntnisse und Perspektiven zur Lösung der jeweiligen Aufgabe bei.

Vielen aktuellen IR Werkzeugen fehlt die Unterstützung für die kollaborativen Aspekte der Informationssuche. Diese fehlende Unterstützung wird in der Praxis durch Behelfslösungen umgangen. Die intensive Benutzung solcher Behelfslösungen zeigt, dass eine Kluft besteht zwischen dem Bedarf der Anwender zusammenzuarbeiten und den Möglichkeiten, die die gegenwärtigen Suchwerkzeuge bereitstellen.

Um diese Kluft zu schließen, wird in dieser Dissertation ein entscheidungstheoretischer Ansatz verfolgt, um optimale Kollaborationsstrategien zu ermitteln und diese den Mitgliedern eines Teams zu empfehlen. Dabei werden die Eigenschaften von Suchaufgaben aus dem Bereich des Patentwesens berücksichtigt, für die bekannt ist, dass sie Recall-orientiert sind.

Basierend auf einer empirischen Benutzerstudie wird ein konzeptuelles Systemmodell entwickelt, welches die technische Umgebung beschreibt, in welcher Kollaboration realisiert wird. Weiterhin wird ein informelles Modell entwickelt, welches den Prozess der Benutzerunterstützung während des kollaborativen IRs beschreibt und welches das konzeptuelle Systemmodell integriert. Anhand der durch die informellen Modelle beschriebenen Rahmenbedingungen wird ein formales Kostenmodell für kollaboratives IR eingeführt. Von diesem Kostenmodell wird ein Ranking-Prinzip abgeleitet, welches das Probabilistische Ranking-Prinzip hin zu Situationen verallgemeinert, in denen mehrere Personen zusammenarbeiten, um ein gemeinsames Informationsbedürfnis zu befriedigen. Weiterhin wird das Konzept der Aktivitätsempfehlungen (engl. Activity Suggestions) entwickelt. Dies ist ein formales Kriterium, welches ebenfalls aus dem Kostenmodell abgeleitet wird und optimale

Kollaborationsstrategien im IR als Lösung eines linearen Optimierungsproblems beschreibt.

Eine prototypische Implementierung zeigt die Umsetzbarkeit von Activity Suggestions. Der Prototyp wurde in einer quantitativen Evaluation eingesetzt, in der die Anwendung von Activity Suggestions anhand der Aufteilung von Suchergebnissen auf die Mitglieder eines Teams demonstriert wird. Dabei werden die Auswirkungen unterschiedlicher Teamgrößen sowie unterschiedlicher Größen von Ergebnissemengen, die je ein Teammitglied bearbeitet, untersucht. Die Resultate zeigen, dass die potentielle Retrieval-Effektivität eines Teams gesteigert werden kann, oder die Aufwände eines Teams, welche benötigt werden, um eine bestimmte Ergebnisgüte zu erreichen, verringert werden können.

## Abstract

In Information Retrieval (IR) research, models and systems traditionally assume that a single person is querying and reviewing the results. However, several empirical studies of professional practice identified collaboration during IR as everyday work patterns in order to solve a shared information need and to benefit from the diverse expertise and experience of the team members. Moreover, most IR systems that are employed in professional work routines are designed for individual use and prototype collaborative systems are too limited to support use in heterogeneous environments of today's work practice.

To bridge this gap, this dissertation develops and formalizes a decision theoretic approach towards supporting a team of people that explicitly set out together to resolve a shared information need. Furthermore, the characteristics of professional search tasks concerning the intellectual property domain are considered. Those tasks are known to be recall-oriented.

Based on an empirical user study, a conceptual system model is proposed that covers the technical environment in which collaborative IR is performed. Moreover, an informal model describing the process of collaborative IR support in such environments and integrating the conceptual system model has been developed. Based on the conditions defined by these informal models, I introduce a formal cost model for collaborative IR and derive a ranking principle for collaborative search session. This ranking principle generalizes the Probability Ranking Principle to situations where several team members work together in a loosely coupled manner and aim at satisfying a shared information need. Moreover, I introduce the notion of Activity Suggestions, that is, a formal criterion that is also derived from the cost model and describes optimum collaboration strategies in IR as the solution of an integer linear program. Those collaboration strategies are suggested to team members with the aim to facilitate the collaborative performance of information retrieval tasks.

A prototypical implementation of Activity Suggestions demonstrates the practicability of the developed formal criterion. The prototype was used for a quantitative evaluation which exemplified the application of Activity Suggestions by means of search result division of query responses among team members. The effects of a changing team size are studied as well as effects of a changing number of documents examined by each team member. The evaluation results show that using the developed criterion, potential retrieval effectiveness is improved or, alternatively, efforts of the team are reduced (in terms of team members required to contribute to the search task) when performing a collaborative, recall-oriented search task.

# Contents

# Figures

# Tables



# Listings

# 1 Introduction

Nowadays, work in modern organizations is to a large extent performed collaboratively. This is particularly true for knowledge work [Blackler 1995]. A result of innovations in communication and information technology is that work teams are distributed across regions and nations. This applies for industry [Fidel et al. 2000] as well as academia [Wilsdon and others 2011].

***Knowledge Work*** is characterized by the handling and distribution of information. Typically, knowledge workers spend a significant amount of their work time searching for information [Sellen et al. 2002]. Collaboration can be a useful strategy in many professional tasks, such as Information Retrieval (IR). The benefit of collaboration results from synergetic interactions between team members, negotiations, discussions and the adoption of other perspectives to produce a common solution or strategy [Taylor-Powell et al. 1998]. Team work exploits the different, often complementary knowledge and abilities of the individual team members [Cummings 2004].

There is ample empirical evidence that in professional practice, IR is often a collaborative process. Studies in engineering [Bruce et al. 2003; Morten Hertzum and Pejtersen 2000; Steven Poltrock et al. 2003], the Intellectual Property (IP) domain [Hansen and Järvelin 2005], legal practice [Attfield et al. 2010], medical care [Reddy and Jansen 2008], academic research [Spence et al. 2005], higher education [Hyldegard 2006; Talja 2002; Twidale et al. 1997], and military [Prekop 2002] highlight the aspect of collaborative activities and show that collaboration within IR processes takes place regularly. These ethnographic field works have identified broad patterns of team and individual behavior related to IR. Collaboration related to IR may include activities such as sharing search strategies, sharing obtained search results, and eventually accumulating relevant information from each team member. Moreover, such collaboration occurs within and across organizational boundaries and ranges from ad-hoc activities to coordinated actions [Böhm et al. 2014b].

## 1.1 Professional Search

In professional practice, searching is part of work duties rather than an activity making sense on its own [Byström and Hansen 2002]. ***Professional Search*** is defined by the fact that the searchers are undertaking the searches on a paid basis which is in contrast to searching on a voluntary basis motivated by personal interest [Tait 2014]. Additionally, professional search is often performed under stringent conditions, e.g., legal requirements [Hunt et al. 2012] [Azzopardi et al. 2010].

Typical examples of professional search can be found in the IP domain. Companies assign teams to survey IP coverage of patents before an investment to figure out which claims already exist and which solutions are free of charge or under license [Landwich, Vogel, et al. 2009]. Also, in a patent office, teams of patent engineers check applications for conflicts with existing patents [Hansen and Järvelin 2005]. Those tasks are information intense and must be exhaustive to avoid patent infringement. However, regarding search practices, the main characteristics can be summarized as follows:

- Professional search is an interactive and iterative process [Bonino et al. 2010] with the aim to gather as much relevant information as possible [Joho et al. 2010]. This means that professional search tasks are often recall-oriented and may be accomplished using several work sessions that can last hours, or even days [Joho et al. 2010]. Typical activities included in such processes are browsing of documents, iteratively revising a query, and analyzing results.
- Professional search is often a collaborative process [Hansen and Järvelin 2005], where searchers aim to benefit from synergetic effects by leveraging diverse sets of knowledge brought in by different people [Shah and González-Ibáñez 2011]. This allows for division of labor and sharing of knowledge among collaborators [Foley and Smeaton 2010]. When teams search with a shared goal, they can benefit from several advantages over individual searching, such as increased coverage of the information space, higher confidence in the quality of their findings, and thus greater productivity [Morris 2007].

However, such professional search activities are limited by budgets [Tait 2014]: The notion of a budget limits the amount of time which may be spent in query formulation and in examining search results: "*All searches are time limited in practice: no one has an infinite amount of time to search for a piece of information*" [Tait 2014].

## 1.2 Collaborative Information Retrieval

The concept ***Collaborative Information Retrieval*** (CIR) is overloaded with several meanings. Most prominent examples are recommender systems, such as the ones provided by online Web shops, e.g., *Amazon.com* [Linden et al. 2003]. This is regarded as implicit collaboration, since people may be generally aware that their results are based in part on past activities of other users, but they may not know who those users were or what information need they had. Thus, collaboration here exists because the search engine used historical data as source of document relevance [J. Pickens et al. 2008]. Explicit collaboration is conceptually distinct by having two or more people who share the same information need and explicitly set out together to satisfy that need [G. Golovchinsky et al. 2008]. Those instances of CIR are ranging from multiple searchers working separately using synchronized single-user interfaces [Morris 2007], or several users sharing one multi-user interface [Smeaton et al. 2007].

The research area CIR lies in the intersection of the computer science fields IR, Human-Computer Interaction (HCI) and Computer Supported Cooperative Work (CSCW). This is schematically depicted in Figure 1.1. HCI research focuses on the interfaces between human and computer [Shneiderman 1992] by observing ways in which humans interact with computers and by designing technologies that let humans interact with computers. IR aims at resolving a human's Anomalous State of Knowledge [Belkin et al. 1982a; Belkin et al. 1982b] using computer-based technology. Research in the area of CSCW has concentrated on computer-based technologies to support people while working together to solve problems. A particular area of interest has been supporting people separated by distance, and helping the establishment of distributed teams that can draw on a wider pool of expertise [Rodden 1991].

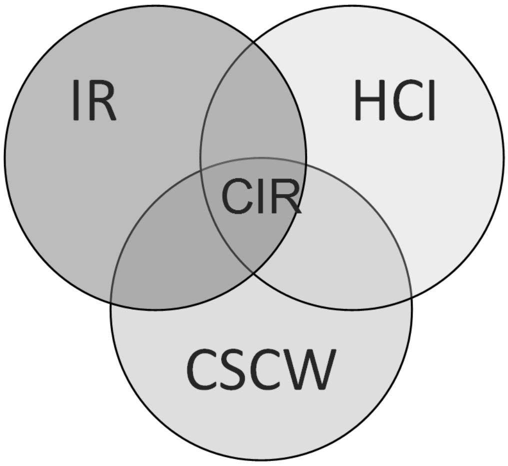



**Figure 1.1:** CIR as research area within the intersection of different fields of computer science

A ***Collaborative Search*** involves multiple users aiming at collaboratively solving a shared information need [G. Golovchinsky et al. 2008]. This sub-discipline of CIR is characterized by explicit collaboration of team members with a shared information need. The objective of the research presented in this dissertation is to support collaborative search activities of focused and geographically distributed teams.

## 1.3 Problem Statement and Motivation

In order to support the different team-based work and information activities in a collaborative environment, a broad set of software tools and systems have been designed. CIR support systems often provide a team with a shared workspace and usually also provide special tools for supporting communication, coordination, and cooperation [Teufel 1995]. Numerous approaches originated in research that allowed team members to collaborate during each stage of the IR process, e.g., (1) query construction, (2) selecting results for inspection, and (3) assessing results [Böhm et al. 2013].

However, recent empirical studies [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013] revealed that, despite the increasing availability of system that are specifically designed to support CIR, in professional practice, search systems and interfaces designed for individual usage are utilized in collaborative work [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013]. This has two main consequences for professional searchers:

First, when people search for information to satisfy a shared information need, they use traditional search engines and interfaces designed for solitary usage. Hence, user support functions implemented by these information systems concentrate on the individual rather than on the team level. The underlying principle of most user support functions is optimizing a list of items towards a particular user. For example, users are provided with ranked lists of documents for inspection or query terms for expansion. Optimum ranking for individuals has been extensively investigated in IR research. According to Gordon and Lenk [Gordon and Lenk 1991] [Gordon and Lenk 1992], user satisfaction is maximized if documents are returned in an order that minimizes the number of documents to be inspected by users to satisfy their information need. The ***Probability Ranking Principle*** (PRP) states that an IR system performs optimally, i.e., cost minimizing, if a list of documents is ranked according to decreasing probabilities of relevance [Robertson 1977]. The PRP has been enhanced by considering IR as an interactive process and by relaxing the assumption of independence between documents [Fuhr 2008]. Furthermore, approaches based on the Portfolio Theory [Zuccon et al. 2010] [Wang and Zhu 2009] and Quantum Theory [Zuccon and Azzopardi 2010] aimed to increase novelty and diversity as well as to cope with interdependent document relevance. However, in all of these approaches, information searchers are still assumed to be individual actors. Little work has been done in developing a general criterion for ranking documents in collaborative sessions, i.e., estimating which document should be inspected by whom.

Second, the above mentioned studies [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013] on CIR practices revealed that team members work independently and that they synchronize their work via loosely coupled communication [Patel and Kalter 1993], i.e., participants use individually preferred applications and use infrequent information exchange to copy the state of work among another. When team members search to satisfy the same information need, they often use the same or very similar query terms [Foley and Smeaton 2009]. Thus, if searching within the same (electronic) information sources, it is likely to result in highly similar ranked lists returned by the search engine. This may lead to less coverage and less productivity due redundant work [Morris 2007].

The aim of this dissertation is to develop a theoretically sound model for ranking documents in collaborative settings. Based on this formal model, the notion of Activity Suggestions for collaborative search is derived. That is a general optimum criterion that is exemplified by search result division. It accounts for the different information activities performed within the team and shifts documents among its members accordingly.

Whereas the PRP is justified by minimizing abstract costs towards an individual searcher [Robertson 1977], *my research hypothesis is that in collaborative search sessions, minimizing abstract costs considering the team as a whole is more effective than minimizing the costs towards each team member individually*.

## 1.4 Research Methodology and Research Questions

Research is commonly understood as a systematic investigation with possible problem solving along with reproducible results. Therefore, suitable research methods are needed to guide the investigation. For the purpose of this dissertation, the framework proposed by Nunameker et al. [Nunamaker Jr and Chen 1990] for information systems research has been chosen as methodological framework. It consists of the following four phases that are depicted as sequence in Figure 1.2. ***Observation*** allows researchers to identify entities involved in the subject of interest and helps formulating specific research hypotheses. It includes methods such as case studies, field studies, and surveys. ***Theory Building*** covers the construction of conceptual frameworks, mathematical models, or the development of new methods. Theories are usually concerned with generic system behaviors. ***Systems Development*** includes prototyping used as a proof-of-concept to demonstrate feasibility. This method formalizes the creation of an artifact as entity to study. ***Experimentation*** includes laboratory or field experiments as well as computer simulations. Experimentation is concerned with the validation of the theories. Results may be used to refine theories or refine systems [Nunamaker Jr and Chen 1990].

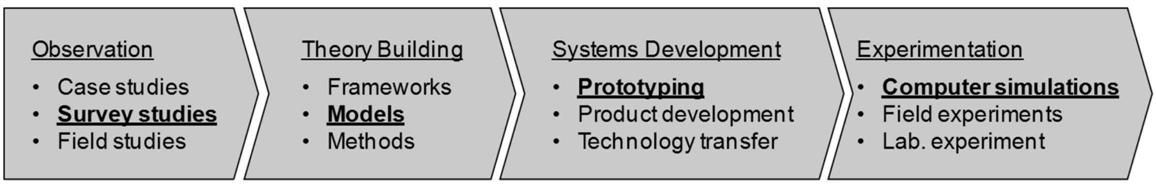



**Figure 1.2:** Research methodology applied in this dissertation **[Nunamaker Jr and Chen 1990]**

The sequence of phases depicted in Figure 1.2 provides a general structure for the research presented in this dissertation. Concrete research methodologies utilized in this dissertation are underlined and highlighted in bold in Figure 1.2. The following concrete research questions are associated with the phases Observation, Theory Building, and Experimentation:

- RQ1 (Observation): *What constitutes a collaborative environment in professional real-world settings used to perform information searching and sharing activities within teams?*
- RQ2 (Theory Building): *How can a team be supported during a collaborative search session? Specifically, how can information search systems be enhanced to reflect team member's information activates?*
- RQ3 (Experimentation): *To which extend does this support increase the potential retrieval effectiveness of the collaborative search tasks?*

## 1.5 Structure of the Dissertation

The dissertation is structured in accordance with the research process illustrated in Figure 1.2. As basis for this research, chapter 2 presents the State-of-the-Art in the area of CIR. Used notions and concepts are introduced and analyzed. In addition, the field of research is concretized and the research gap which represents the motivation for this dissertation is identified.

- Observation*:* Section 3.1 reports on an empirical user study that has been conducted to investigate the CIR practices of three work groups in academic and industrial research facilities.
- Theory Building*:* In sections 3.2, the findings of the user study are captured in a conceptual system model that covers the technical environment in which CIR is performed. Moreover, an informal model describing the process of CIR support in such environments and integrating the conceptual system model has been developed. In section 3.3, a probabilistic model describing CIR activities and its associated (abstract) costs is introduced. From this cost model, a ranking principle for collaborative search session is derived. Finally, in section 3.4, Activity Suggestions are developed, i.e., a formal criterion (that is also derived from the cost model) that describes optimum collaboration strategies in IR as the solution of an integer linear program.

- Systems Development*:* Chapter 4 provides a brief description of a prototype implementation of Activity Suggestions as developed in section 3.4.
- Experimentation*:* Chapter 5 presents a thorough evaluation of Activity Suggestions based on a computer simulation of collaborative search tasks in two professional domains. The evaluation methodology is described in section 5.1. Evaluation results and significance tests are presented in section 5.2.

Chapter 6 closes this dissertation with summary and concussions as well as an outlook on future research.

## 1.6 Contributions

The major scientific contributions of this dissertation can be summarized as follows:

- This dissertation introduces a novel ranking principle, denoted with cPRP, along with its proof of optimality. It generalizes the PRP to situations where multiple users work together in a loosely coupled manner and aim at satisfying a shared information need.
- This dissertation introduces the notion of collaborative Activity Suggestions, i.e., a formal criterion that describes optimum collaboration strategies in IR as integer linear program.
- New insights on CIR practices of professionals in academic and industrial research facilities are described. A conducted user study provided insights on Information Technology (IT) equipment utilization in professional work used for realizing collaboration. The study also has been published in [Böhm et al. 2013] and [Böhm et al. 2014a].
- A summarizing model describing Social and Work Tasks Aspects in CIR is presented. It describes CIR as central pattern at all levels of a work task, within and across organizational boundaries. This model has also been published in [Böhm et al. 2014b]

- A classification of collaborative information activities at IR level is presented. This classification has also been published in [Böhm et al. 2013] and [Böhm et al. 2014a].

## 1.7 Publications

The following publications have arisen as part of this dissertation:

1. Böhm, Thilo, Claus-Peter Klas, and Matthias Hemmje. "An Experimental Evaluation of Collaborative Search Result Division Strategies", *19th International Conference on Theory and Practice of Digital Libraries,* Pozan, Poland, 2015.
2. Böhm, Thilo, Claus-Peter Klas, and Matthias Hemmje. "A Probabilistic Ranking Principle for Collaborative Search", *Collaborative IS: Best Practices, New Insights, New Thoughts;* Eds.: Hansen, Shah, Klas. Springer, 2015.
3. Böhm, Thilo, Claus-Peter Klas, and Matthias Hemmje. "Collaborative IS and Retrieval in a Heterogeneous Environment." *Computer* 47.3 (2014): 32-37.
4. Böhm, Thilo, Claus-Peter Klas, and Matthias Hemmje. "Collaborative IS in Professional Work-Settings: A Study of Equipment Utilization." *Datenbank-Spektrum* 14.1 (2014): 29-38.
5. Böhm, Thilo, Claus-Peter Klas, and Matthias Hemmje. "Supporting Collaborative IS and Searching in Distributed Environments." *Proceedings of the LWA 2013 conference*, 2013.
6. Boehm, Thilo. "Group-support for task-based information searching: a knowledge-based approach." *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013.

# 2 State of the Art

This chapter introduces used notions and concepts. It presents and analyzes previous fundamental work in the area of CIR. Additionally, the field of research is concretized. This is done in the following order.

In order to provide an understanding of underlying concepts related to the fields covered in this dissertation, section 2.1 introduces the most important (informal) definitions.

The objective of this dissertation is to support a team of searchers during their information activities. To understand the various components involved in such activities and to identify factors are likely to have an influence, several empirical field works aimed at capturing collaborative activities of work teams in real-world settings which I review and summarize in section 2.2 and 2.3.

In section 2.4, I provide an analysis of the various kinds of computer-based technology developed to support collaboration in IR. This section is structured according to the sub-fields of CIR in accordance to a taxonomy introduced in [G. Golovchinsky et al. 2008].

Because the main focus of this dissertation is to support explicit collaboration in IR, section 2.5 reviews research that was particularly concerned with this sub-field of CIR. In section 2.5, I develop a schema for classifying software tools aiming to support collaborative information activities. The reviewed research is classified in accordance to the developed schema.

The final stage of the research methodology employed in this dissertation is the evaluation of the developed entity. In section 2.6, I discuss the problems with IR evaluation and describe the methodologies used in standard IR evaluation and how these methodologies were adopted towards CIR evaluation.

Finally, the State-of-the-Art is summarized and discussed in section 2.7. Moreover, the research gap is identified which represents the research objective and motivation of this dissertation.

## 2.1 Basic Concepts

This section introduces a set of concept definitions selected from literature. Firstly, this covers the fields of Information Science and Information Retrieval, since one objective of this dissertation is to support searchers in their endeavor to find required information (see sections 2.1.1 and 2.1.2). Because the related activities typically focus on a specific problem or task, corresponding definitions are summarized in section 2.1.3. Secondly, concepts from the field of CSCW are introduced because I focus on information searchers that work together when resolving their information need (see section 2.1.4 to 2.1.7).

### 2.1.1 Information and Information System

Concepts such as ***Data***, ***Information***, and ***Knowledge*** as well as their relationships were often defined based on a hierarchal understanding. This means that data was seen as the basis, information was defined in terms of data, and knowledge was defined in terms of information [Rowley 2007]. In more detail, data is considered discrete facts about objects or events. Information is the result of analysis, processing, or classifying of data. Knowledge is considered the distillation of information to incorporate experience, values, insight and intuition [Davenport and Prusak 1998]. However, Information Science provides a more pragmatic definition of these concepts which will be summarized below.

Kuhlen deduced the concept of information from that of knowledge in such a way that *"information is the sub-set of knowledge which is needed by but not available to a specific person in a concrete situation in order to solve a problem"* [Kuhlen 1991]. This definition sees information as not existing in itself. Information is knowledge that becomes effective and relevant with regard to a specific purpose. Knowledge in order to become information, must be represented, either using a natural language or a knowledge-representation language. Information references not only to represented knowledge, but unfolds its meaning only with reference to a particular use [Kuhlen 2004].

Kuhlen defines data as measured units which are obtained through observation of natural, constructed, or simulated objects and events. Data is stored and represented using syntactically well-defined rules in an agreed system of symbols. However, data becomes information if it is selectively retrieved from an information system in a particular context and/or with respect to a decisive purpose [Kuhlen 2004]. An ***Information System*** is a computer-based system providing means for the collection, organization, storage and communication of information [Bates 2011]. However, according to Kuhlen this is a terminological inconsistency because such systems do not contain information as long as there is no inquiry to the system and the response is not utilized by anyone [Kuhlen 1991]. However, Kuhlen argues that data collected in an information system has the potential to become information [Kuhlen 2004]. The user activities supported by an information system and its components are called ***Information Activities*** [Hemmje et al. 1996]. The data stored in an information system that is access-able via information activities are called ***Information Objects*** [Belkin et al. 1995]. Information systems that operate on mostly unstructured data, such as text documents, are called ***Information Retrieval Systems*** [Bates 2011]. The purpose of an IR system is to help a user satisfying his or her information need.

### 2.1.2 Information Need and Information Behavior

An ***Information Need*** arises when a human encounters an ***Anomaly State of Knowledge*** (ASK) [Belkin et al. 1982a; Belkin et al. 1982b]. Belkin described the ASK as a situation where *"the user realizes that there is an anomaly in [their] state of knowledge with respect to the problem faced"*. The user's recognition of insufficient knowledge triggers activities to locate and obtain information required solve the problem. More generally, those activities are summarized by the concept of information behavior.

Wilson defined ***Information Behavior*** (IB) as *"the totality of human behavior in relation to sources and channels of information"* [Wilson 1999]. This involves the generation, acquisition, use, and communication of information. For example, one concrete IB would be that users address their recognized ASKs by searching for information which eventually leads to the use of an IR system. Wilson furthermore identified nested sub-levels. Those are ***Information Seeking*** (IS) which generally focuses on the variety of methods people employ to discover and gain access to information sources, and ***Information Searching*** which is particularly concerned with the interactions between information user and computer-based information systems. Wilson's definition incorporates previous models of IS, such as Dervin's sense making model [Dervin and

Nilan 1986], Ellis' behavioral model of IS [Ellis 1989], and Kuhlthau's ***Information Search Process*** (ISP) [Kuhlthau 1991].

### 2.1.3 Work-Tasks

Information-intensive work in professional settings involves dynamic information utilization in which IR tasks are taking a central role. Marchionini described the generic term ***Task*** as the manifestation of a user's problem. It is the task that drives IR activities [Marchionini 1997]. Byström and Hansen described a task as an activity that is carried out to achieve a specific goal [Byström and Hansen 2005]. However, the concept of a ***Work Task*** represents a specific task that is carried out by individuals to fulfill their work duties. Work tasks are situated in the work organization and reflect organizational, cultural and social norms, as well as organizational resources and constraints [Byström and Hansen 2005].

According to Byström and Hansen, a work task may also be seen as a process wherein an actor performs a set of actions with a specific item of work in focus. A work task may consist of several sub-tasks that may further consist of information activities, such as query formulation or relevance assessment. As a result of an identified information need, a work task may include IS sub-tasks that are further decomposed into IR sub-tasks (or information searching tasks, respectively) [Byström and Hansen 2005]. ***Information Seeking Tasks*** generally focus on the satisfaction of complex information needs and involve several sources and consultations of them. An ***Information Retrieval Task*** is particularly concerned with the satisfaction of a separable fraction of an information need through consultation of one electronic source, such as a specific digital library. Consulting a human source is an example of a single ***Information Searching Task***.

### 2.1.4 Collaboration

Although literature does not provide a clear definition of the concept of collaboration, the references summarized here provide insights into the main characteristics of collaboration. Moreover, concepts such as collaboration and cooperation are often used synonymously. Therefore, I introduce a set of (informal) definitions of related concepts as defined in [Shah 2011].

***Collaboration*** is generally understood as "*to work together*" [London 1995]. Group work exploits the different, often complementary knowledge and abilities of the individual group members [Cummings 2004] who work towards a common purpose by means of communication, interactions, information sharing, and coordination of activities [Amabile et al. 2001; Melin and Persson 1996]. Denning and Yaholkovsky suggested that coordination and cooperation are weaker forms of working together and that all of these activities require sharing some information with each other [Denning and Yaholkovsky 2008]. Effective collaboration requires each member of the group making an individual contribution [Taylor-Powell et al. 1998]. The benefit of collaboration results from synergetic interactions between individuals, negotiations, discussions, and the adoption of other perspectives to produce a solution or strategy which incorporates the different backgrounds of the collaborators [Taylor-Powell et al. 1998].

In an attempt to capture the different facets of collaboration, Shah synthesized a model of collaboration consisting of five different layers [Shah 2011]. This model is depicted in Figure 2.1.

***Communication*** represents the center of the model and is seen as a process of exchanging information. This is considered as the most basic requirement for the development of collaboration. ***Contribution*** covers activities through which participants of the group work support each other to achieve their goals. Contribution is viewed as an informal rather than a formal relationship. **C*oordination*** is seen as a process of connecting individuals that may share resources, responsibilities, and goals. ***Cooperation*** covers planning activities, discussing and negotiating roles as well as sharing resources among each other. It also assumes that the participants follow a set of common rules which represents a further difference to coordination. Finally, collaboration is the process where people work together. They may see different aspects of a problem and contribute with their personal expertise towards accomplishing the group work.

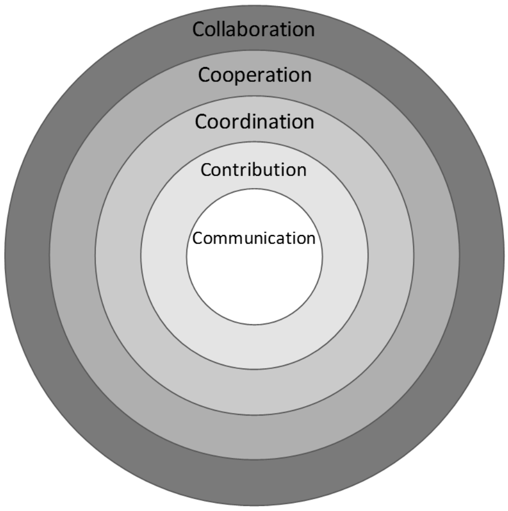



**Figure 2.1:** Five-layer model of collaboration **[Shah 2011]**

### 2.1.5 Groups and Roles

The collaborative performance of a work task is being conducted by a team which represents a specialization of the more generic term group. A ***Group*** can be understood as a collection of interacting individuals [McGrath et al. 1993]. According to McGrath et al., it consists of at least two people that are mutual aware of one another and that potentially interact with each other. With respect to collaboration within groups, three specializations are commonly used in literature: team, social network, and community [Rohde and Shaffer 2003]. ***Teams*** are small groups characterized by a common task they need to complete. Their members are often called ***Co-workers***. Teams are embedded in organizations and sub-organizations, respectively. ***Social Networks*** are defined by links between its members and those links represent social contacts. These social contacts indicate the network's structure. A ***Community*** is a group which's individuals are characterized by a shared culture, i.e., *"a set of shared norms, conventions, and meanings; a set of common practices and common symbols that create a shared semantic space"* [Rohde and Shaffer 2003].

### 2.1.6 Collaboration Support Systems

The term CSCW was coined by Cashman and Greif for a workshop in 1984 [Grudin 1994]. Research in the area of CSCW has concentrated on technologies to support people working together to solve problems. "*In its most general form, CSCW* *examines the possibilities and effects of technological support for humans involved in collaborative group communication and work processes*" [Bowers and Benford 1990].

A particular area of interest has been supporting people separated by distance, helping the establishment of distributed teams that can draw on a wider pool of expertise [Rodden 1991]. Computer systems designed to facilitate action-oriented teams working together over geographic distances have been phrased ***Groupware*** [Ellis et al. 1991]. Literature provides different approaches of classifying various kinds of Groupware systems. Teufel et al. classified systems based on the 3-C model, i.e., according to their degree of supporting communication, coordination, and cooperation between group members [Teufel 1995]. Figure 2.2 illustrates this classification. It can be seen that the position within the triangle expresses the degree of supporting the three phenomena.

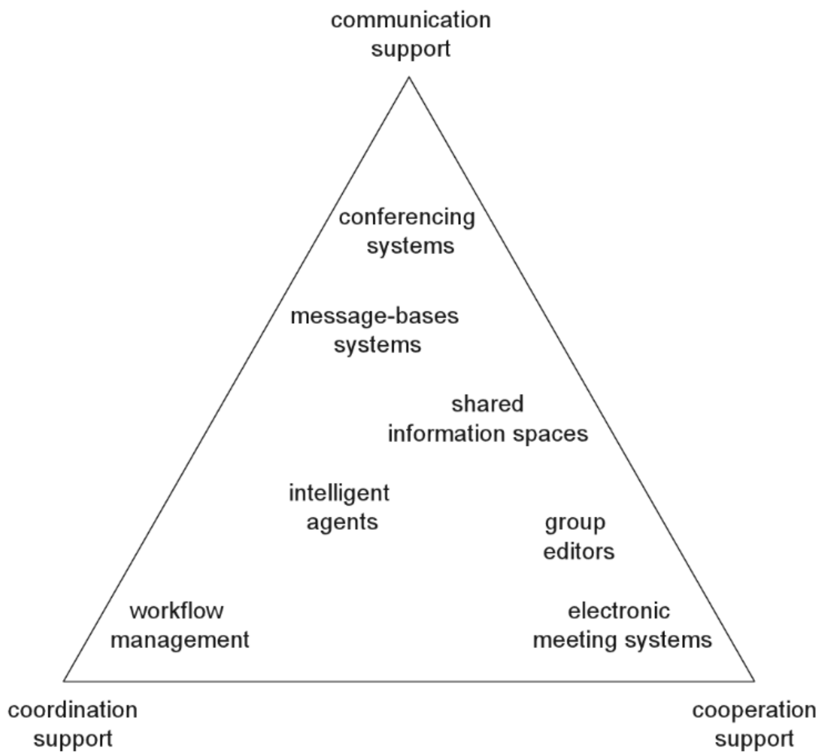



**Figure 2.2:** Classification based on the 3C model **[Teufel 1995]**

Another way of classifying groupware systems is by determining the place and time of the collaborative interactions that are being supported [Johansen 1988]. Collaboration, as depicted in Figure 2.3, may occur at the same time or separated in time (***Synchronous*** vs. ***Asynchronous***), and may occur at the same place or separated by space (***Co-located*** vs. ***Remote***). Examples from the various quadrants are:

- Same time, same place: meeting support tools,
- Same time, different place: video conferencing,
- Different time, different place: e-mail systems,
- Different time, same place: corporate workflow systems running over an intranet.

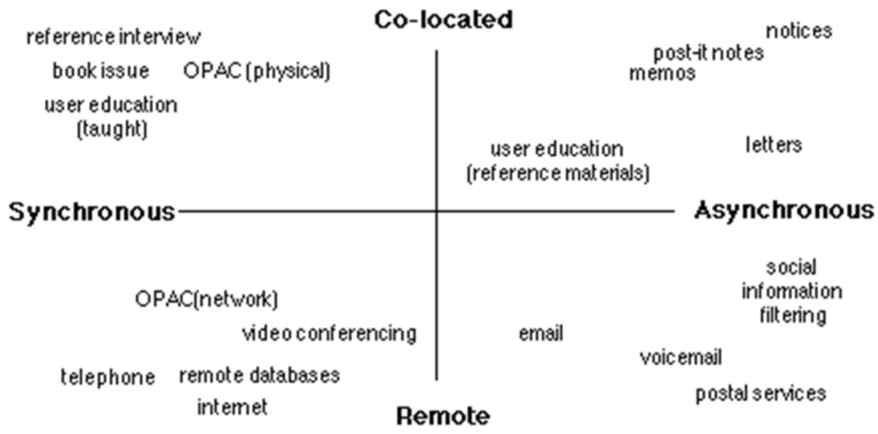



**Figure 2.3:** System classification based on the space/time matrix **[Twidale and Nichols 1998]**

In addition, there are a number of application classes of systems that have been developed to support different kinds of collaborative interactions. Surveys and classifications of systems have been conducted, for example, in [Ellis et al. 1991] as well as in [Grudin 1994]. Finally, Twidale and Nicols did a survey on CSCW applications in the field of digital libraries [Twidale and Nichols 1998]. Figure 2.3 depicts the quadrants of the time/space matrix using actual and potential collaborative activities in a library context, as identified in [Twidale and Nichols 1998].

### 2.1.7 Information and Knowledge-Sharing in Organizations

If multiple individuals of an organization make use of knowledge to advance organizational goals (e.g. accomplish work tasks), such an organization is called a ***Knowledge-centric Organization*** [Crawford et al. 2009]. Knowledge-centric organizations collect and leverage heterogeneous sources of data and knowledge and make them a core value. Its members are called ***Knowledge Workers*** that produce ideas and information.

Nonaka and Takeuchi created a theoretical description of knowledge creation in organizations [Nonaka and Takeuchi 1995]. Nonaka and Takeuchi theorized that the creation of knowledge is the result of a continuous cycle of four integrated processes: (1) externalization, (2) combination, (3) internalization and (4) socialization. These four knowledge conversion mechanisms are the combinations of conversions of explicit and tacit knowledge.

Within a team aiming at work task completion, individuals share (implicitly and explicitly) knowledge and information [Sonnenwald et al. 2004]. Wilson and Järvelin identified three classes of knowledge required for work task performance [Järvelin and Wilson 2003]. Those are:

1. ***Problem Knowledge*** which describes the structure, properties and requirements of the problem at hand,
2. ***Domain Knowledge*** that consists of known facts, concepts and theories in the domain of the problem, and
3. ***Problem-solving Knowledge*** which describes how problems should be treated and what problem and domain knowledge should be used to solve the problem.

For example, in the context of an IR task, the problem-solving knowledge is constituted by search skills of an individual, e.g., search strategies and search tactics as described in [Bates 1979a; Bates 1979b]. For example, Bates identified ***Search Strategies*** of information system users as *"A plan for the whole search",* whereas ***Search Tactics*** are considered *"A move made to further a search".* An example for the latter one would be replacing query terms with more general or specific ones.

Different degrees of domain knowledge and problem-solving knowledge may lead to teams that consist of different combinations of domain experts and novices or (with respect to IR tasks) search expert and novices. Such team structures have been denoted with ***Asymmetric Teams*** [Golovchinsky et al. 2009]. Figure 2.4 depicts this schematically: For a given work task, knowledge covered by two team members is visualized exemplary.

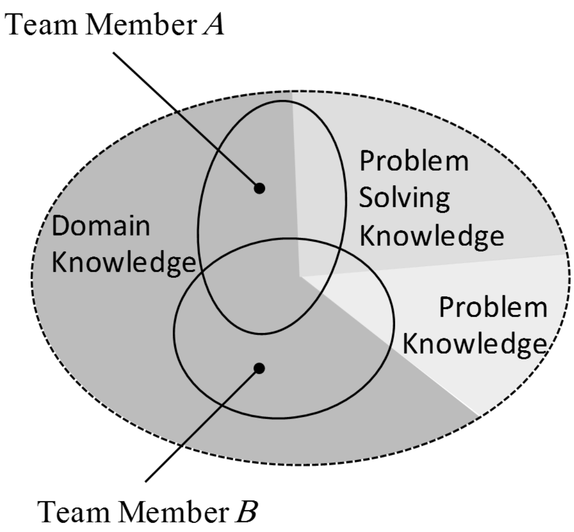



**Figure 2.4:** Schematic depiction of an asymmetric team of two

## 2.2 Collaboration in Information Behavior

Wilson's model of IB (see section 2.1.2) has been revised over time [Wilson 1999] and (in its latest version) aims at linking and integrating other models of IS [Dervin and Nilan 1986], [Ellis 1989], [Kuhlthau 1991], as well as information searching [Belkin et al. 1995] [Saracevic 1997] [Spink et al. 1998]. The resulting problem-solving model by Wilson described a person moving from uncertainty to increasing certainty via the following stages [Wilson 1999]: Problem Identification (where the person is asking the question, "*What kind of problem do I have?"*), Problem Definition (*"Exactly what is the nature of my problem?"*), Problem Resolution *("How do I find the answer to my problem?"*) and, potentially, Solution Statement (*"This is the answer to the problem."*).

Wilson's problem-solving model views IB primarily from a single user's perspective. This is especially outlined by person's questions associated with stages, such as *"How do **I** find the answer to **my** problem?"* [Wilson 1999]. However, early studies of IB indicate the existence of collaborative aspects during IS and information searching:

- Bates identified search tactics of information system users that also included social elements [Bates 1979a; Bates 1979b]; Table 2.1 summarizes these search tactics.

- Kuhlthau's ISP indicated collaborative elements in early stages of the ISP where *„typical actions are to confer with others*“ [Kuhlthau 1991].

**Table 2.1:** Sub-set of Bates' tactics with social aspects **[Twidale et al. 1997]**

| Tactic Name | Description of Tactic |
|---|---|
| CONSULT<br>Idea tactic | To *consult* is described as asking a colleague for help |
| WANDER<br>Idea tactic | To *wander* can be viewed in a similar way to *consult*, as the resources available to a searcher are not limited to physical items but can include people and computerized systems. |
| BRAINSTORM<br>Idea tactic | To *bibble* is to take advantage of searches that have been done in the past and not waste time and resources re-inventing the wheel. A bibliography is a structured version of the results of a past search. |
| BIBBLE<br>Information search tactic | *Brainstorming* can be a personal activity but is more commonly used by groups. |

With the aim of analyzing collaboration in IB with respect to work task completion, several studies have been conducted in many domains of work, such as engineering [Bruce et al. 2003; Morten Hertzum and Pejtersen 2000; Steven Poltrock et al. 2003], the IP domain [Hansen and Järvelin 2005], legal practice [Attfield et al. 2010], medical care [Reddy and Jansen 2008], academic research [Spence et al. 2005], higher education [Hyldegard 2006; Talja 2002; Twidale et al. 1997], and military [Prekop 2002]. These studies identified a variety of collaborative methods and practices people employ to discover and gain access to information. They show that CIR is a common pattern in everyday work. These studies also show that these collaborative activities not only belong to the IS level but also to the IR level [Hansen and Järvelin 2005]. However, any IR activity (such as formulating queries, examining results, and viewing documents) may be performed by an individual on behalf of a team or in collaboration [Steven Poltrock et al. 2003].

In an attempt to integrate the various terminologies associated with ***Collaborative Information Behavior*** (CIB), Karunakaran et al. provided a working definition of CIB as the “*totality of behavior exhibited when people work together to identify an information need, retrieve, seek and share information, evaluate, synthesize and make sense of the found information, and then utilize the found information*” [Karunakaran et al. 2010]. Talja and Hansen see CIB with respect to the accomplishment of an underlying work task and define it as *"... an activity where two or more actors communicate to identify information for accomplishing a task or solving a problem"* [Talja and Hansen 2006]. Moreover, Karunakaran et al. identified ***Collaborative Information Seeking*** (CIS) as a nested field of CIB and defined it as the purposive seeking of information by two or more individuals.

Shah developed a summarizing depiction of collaboration and IB [Shah 2011]. Figure 2.5 extends his schematic depiction by adding the concept of CIB as defined above. Figure 2.5 depicts IR as a subset of IS, both nested as a specific IB. Shah places CIS and CIR in the context of collaboration. Collaboration includes several parts, some of which may not be related to IS and IR, such as coordination of activities.

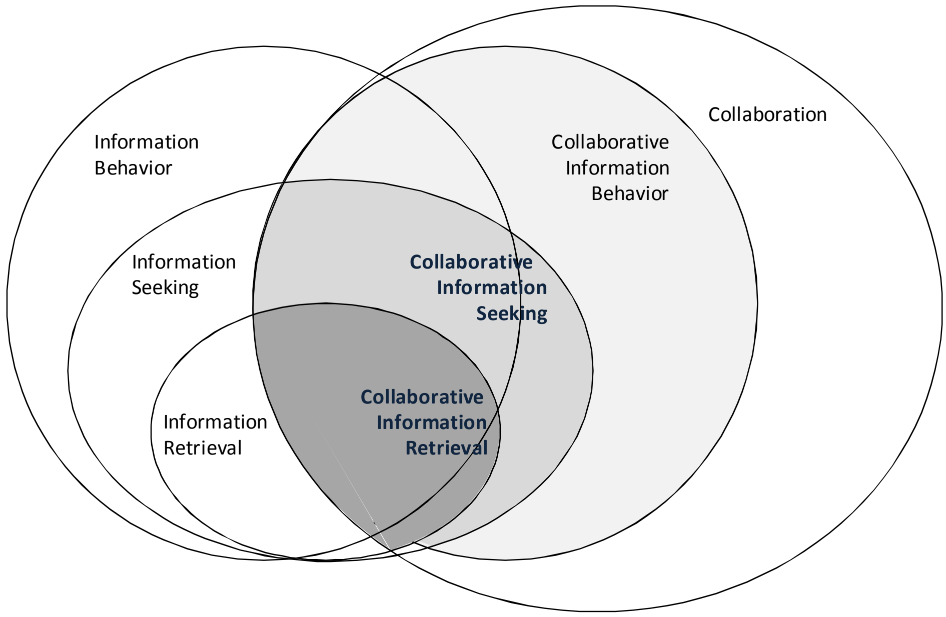



**Figure 2.5:** Schematic depiction of nested fields related to the definition of CIB **[Shah 2011]**

## 2.3 Collaboration at Information Seeking Level

As it has been pointed out in section 2.2, early empirical studies of IB observed several collaborative aspects related to IB. To obtain a better understanding of those aspects, several empirical studies have been conducted to capture collaborative activities related to IS. This was done by observing team members in different real world settings, such as academic research (see section 2.3.1) and industry (see section 2.3.2). Besides these behavioral studies, in section 2.3.3, field works towards identification of IT equipment usage in professional contexts are summarized. Finally, section 2.3.4 presents conceptual models that aimed to incorporate findings and observations gathered using the empirical studies presented here.

### 2.3.1 Academic Research

In [Hyldegard 2006; Hyldegard 2009], Hyldegard investigated, how well Kuhlthau's ISP [Kuhlthau 1991] model explained CIS activities of students in educational settings. She found that the behavior of individual group members was, to some extent, similar to that described in the ISP model. However, group member's behavior also differed in many ways from the ISP model. These differences were found to be related to contextual, social and personal factors. Hyldegard concluded that the ISP model did not completely meet the social dimension of CIS and proposed the Group Member in Context (GMIC) model as an extension to the ISP. In addition, Shah and Gonzáles-Ibánez attempted to map Kuhlthau's ISP model to CIS [Shah and González-Ibáñez 2010]. They discovered that exploration, formulation, and collection were not distinct stages. Participants went back and forth between these stages by trying search queries, exploring various sources, and collecting relevant information as they worked through the task while interacting with their collaborators. In addition to Hyldegard's finding that ISP model lacks social elements in a collaborative setting, the work reported by Shah and Gonzáles-Ibánez indicates that various ISP stages in CIS settings also need to be considered in the light of affective dimension for the collaborators as well as the group's affective relevance.

Sonnenwald et al. investigated the types of information and knowledge that need to be shared to support situational awareness [Sonnenwald et al. 2004]. The authors also studied the ways in which technology can be used to facilitate such sharing of information and knowledge. Their main finding was the identification of three types of information that each team member required to support his or her situational awareness: "*Researchers discovered that successful scientific collaboration requires the collection and use of a range of awareness information that updates team members on the current state of their teams' activities*". The three types of awareness information were contextual-, task-and-process-, and socio-emotional information. Moreover, technology was identified as an important element in mediating these awareness information.

CIS activities often involve sharing of information which is characterized by sharing acquired information among collaborators. ***Information Sharing*** denotes direct information exchanges among those involved in solving a problem. Such information exchanges qualify as CIS only, if actors collaborate to acquire information they did not already have [Fidel et al. 2000]. Talja observed and classified different types of information sharing in an academic environment [Talja 2002]. These types were (1) strategic sharing, (2) paradigmatic sharing, (3) directive sharing, (4) social sharing, and (5) no sharing. Her investigations revealed that in academia, CIS is as common as individual IS. Scholars usually belong to different social networks. According to Talja, these social networks not only influence their choices of IS, but are the place where information is sought, interpreted, used, and created.

### 2.3.2 Industry

Hertzum and Pejtersen conducted two case studies involving engineers [Morten Hertzum and Pejtersen 2000]. They reported on CIS activities and on the importance of providing support for individuals when accessing information systems. They found that IS often involves looking for informing documents as well as looking for informed people. Accordingly, people search for documents to find people, search for people to get documents, and interact collaboratively and socially to acquire information without approaching any explicit search activity. Hertzum and Pejtersen concluded that individuals are a critical source of information because they have experience and knowledge about the context of, e.g., documents that cannot be derived from the content alone. Therefore, they argued that it is necessary to consider the need to consult people with specific competencies and experiences.

In [Bruce et al. 2003] and [S. Poltrock et al. 2003], a study was presented that captured collaboration during information activities of members of two design teams in technology firms when collaboratively performing IS and IR tasks. They observed that the team members actually did collaborate, e.g., when creating search strategies for solving a problem. The authors found that CIR is an integral part of the daily work to solve shared information needs of the team. Identifying, analyzing, and defining the information need, as well as the development of search strategies was performed collaboratively. This involved intra-team as well as extra-team collaboration [S. Poltrock et al. 2003].

Morris conducted a survey regarding Web-search practices among the employees of a large IT company [Morris 2008]. She found that collaboration was commonplace and the majority of all users reported to have used some sort of collaboration when searching the Web. Morris reported that respondents performed activities regarding both, the ***Search Product*** (i.e. useful links, information found within Web-sites) and the ***Search Process*** (i.e. search terms, search queries).

Prekop's qualitative study investigated the collaborative perspective of IS in the military domain [Prekop 2002]. The study described the CIS activities of a work team created to perform command control support. Three components of this collaboration were identified: (1) Information Seeking Roles (ISRs), (2) Information Seeking Patterns, and (3) the Context in which the roles and patterns were performed. ISRs were both, formally assigned and informally adopted. Group members also acted in several ISRs, and several members potentially fulfill the same ISR. Those ISRs were, for example, the Information Gatherer, who's main task was to find information, and the Information Seeking Instigator, who's main task was to direct participants to gather specific information and to initiate the IS. Additionally, supporting roles were described such as the Group Administrator and the Group Manager. Prekop also identified patterns of interaction between ISRs that describe sequences of actions, interactions, and behaviors performed by participants towards accomplishing a work task.

Collaborative activities have been studied and observed also among search intermediaries [O'Day and Jeffries 1993]. O'Day and Jeffries observed and identified the following collaborative strategies: (i) Sharing search results with other members of a team, (ii) self-initiated broadcasting of interesting information, (iii) acting as a consultant and handling search requests made by others, and (iv) archiving potentially useful information into group repositories. O'Day and Jeffries proposed four types of sharing of information in collaborative group situations: (1) Sharing results with other team members, (2) Self-initiated dissemination and broadcasting of interesting information, (3) Using other people's search requests, and (4) Storing potentially useful information in repositories for others to use.

A study conducted by Hansen and Järvelin analyzed the IB of the employees of the Swedish patent office when engaged in the patent application process [Hansen and Järvelin 2005]. They observed collaborative activities in all stages of the IS process which they have categorized as follows: (i) Planning tasks, (ii) Problem definition, (iii) Search topic selection, (iv) Query construction, and (v) Relevance assessments. The authors categorized the observed collaborative activities into document-related and human-related activities. Their study showed that collaborative activities are an important characteristic of IS tasks in professional settings. Hansen and Järvelin identified two new classes of sharing of information in addition to the ones identified by O'Day and Jeffries: (5) Case-building and (6) History-building.

Reddy and Spence presented a field study regarding CIS in multi-disciplinary teams in the context of medical care [Reddy and Spence 2008]. The authors identified four triggers of CIS activities: (1) Complexity of the information need, (2) Fragmented information resources, (3) Lack of domain expertise, and (4) Lack of immediately accessible information.

### 2.3.3 Studies of Equipment Usage in Professional Settings

CIS was also investigated from a more technical point of view. Twidale et al. observed collaboration between students at the computer terminals of the university library, although these systems were not designed for collaborative usage [Twidale et al. 1997]. The authors identified several collaborative activities, such as asking for help, re-using searches, as well as joint and coordinated searching. They categorized the observed activities into search product-related and search process-related activities. Morris conducted a survey regarding Web-search practices among the employees of a large IT company [Morris 2008]. Similar to Twidale et al., Morris identified activities regarding the search product and the search process.

Common to both studies is that multiple people were engaged in IS and IR activities and combined their efforts in pursuit of a common, or at least similar, information need. People communicated about the search process and the search products, but neither the user interface nor the search engine (or more generally the information system) were aware that people intended to collaborate. Other activities of the search process took place outside the system that managed the communication. These studies investigated situations when the collaboration was mediated by face-to-face communication and computer-based applications. The latter included the use of generic communication and data exchange applications, such as e-mail.

The need to collaborate during IS has been addressed by many systems. Corresponding support systems have been developed, to a large extent, in experimental settings and often provided participants with an environment that included shared workspaces and usually also included special collaboration mechanisms. Those means allowed multiple searchers working separately in parallel with shared interface awareness [Morris 2007], or multiple users sharing a single interface [Smeaton et al. 2007].

However, recent empirical studies investigated utilization of IT equipment for realizing collaboration while searching [Morris 2013] [Kelly and Payne 2014]. They found that, despite the increasing availability of systems that are specifically designed to support CIS, these technologies were not used in practice. Instead simpler communications technologies, that are part of the everyday work, were used as means to realize CIS. Team members participating in a collaborative project may find themselves in different locations or settings and may use a variety of different software applications.

In [Morris 2013], Morris followed on her previous study on Web-search practices [Morris 2008] and compared earlier results with her recent survey outcome. The surveys indicated a need for features that support persistence, awareness, and division of labor. Morris found that today, more people are engaging in collaborative Web-search and that they are doing so with higher frequency. The study showed that the increased prevalence of collaborative Web-search is the result of a change in the technology landscape, i.e., today's importance of social networking sites and the use of smart phones. Kelly and Payne support Morris' recent findings that CIS solutions must be low-effort and "*sufficiently lightweight compared with status quo ad hoc solutions*" [Kelly and Payne 2014]. They further suggest that future solutions could be scaled back in favor of lightweight support for core CIS behaviors. They report that CIS support could be embedded in larger applications that support a broader range of high-level planning tasks.

### 2.3.4 Summarizing Models

Based on the various studies in the field of CIS, few conceptual models have been synthesized to summarize earlier work.

A ***Conceptual Model*** provides a working strategy, a scheme containing general, major concepts and their interrelations. It orients research towards specific sets of research questions and forms the basis of formulating empirically testable research questions and hypotheses [Järvelin and Wilson 2003]. As a special kind of a conceptual model, ***Summary Models*** provide overviews of research domains and list factors affecting a phenomena [Järvelin and Wilson 2003]. In the remainder of this section, I briefly outline two summarizing models from the research area of CIS.

#### 2.3.4.1 Model of Collaborative Information Behavior

Karunakaran et al. presented a three-phase model of CIB with a focus on conceptualizing the broad set of activities that could potentially occur within CIB [Karunakaran et al. 2010]. The first phase starts with the process of Problem Identification where a group identifies a problem and creates a shared representation collaboratively. Triggers act as critical transition points from Individual Information Behavior (IIB) to CIB. The four major triggers have been identified in [Reddy and Spence 2008] and are listed in section 2.3.2.

The second phase of activities within CIB gets triggered when the Problem-in-Context meets the set of characteristics (see section 2.3.2), and gives rise to CIS. Here, CIS is understood as the purposive seeking of information by two or more individuals. It may involve the use of a variety of systems, people, and channels in order to address the information need. Karunakaran et al. conceptualized that CIS is comprised of other activities such as retrieving and sharing.

In the third phase (Information Use), the gained information is used to resolve the problem in the context of a given work task. The use of information includes physical, mental, and communicative acts involved in incorporating the information found into the group's existing knowledge base. These three phases are summarized graphically in Figure 2.6.

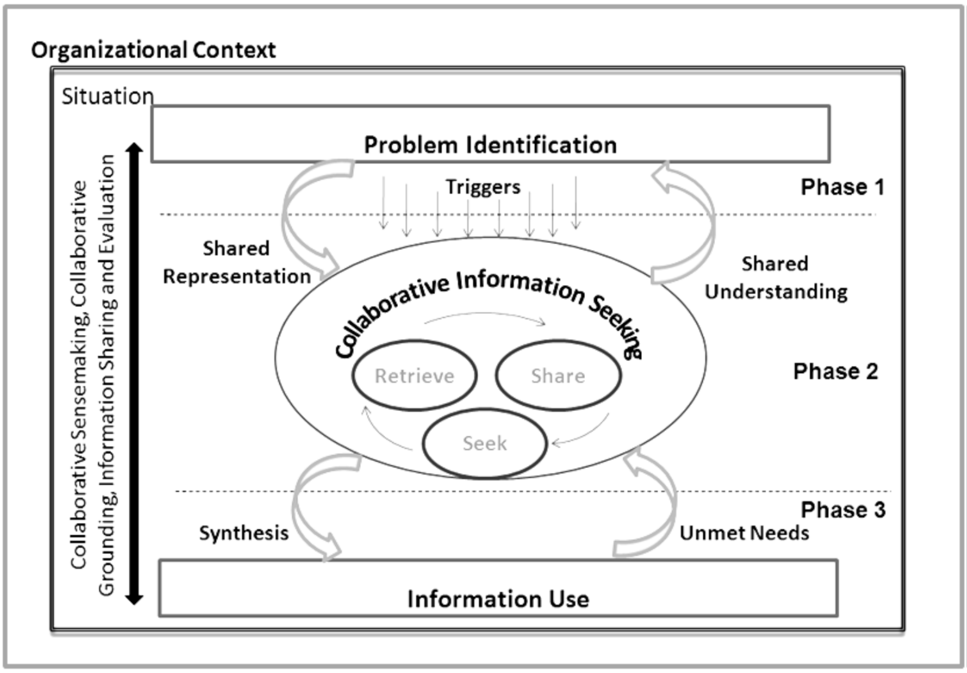



**Figure 2.6:** Three-phase model of CIB **[Karunakaran et al. 2010]**

2.3.4.2 Model for Social and Work Task Aspects in CIS

Based on several empirical studies (see sections 2.3.2 and 2.3.3), I synthesized an model called ***Social and Work Task Aspects in CIS*** [Böhm et al. 2014b] that informally describes collaboration as central pattern at all levels of a work task and across organizational boundaries. This summarizing model (depicted in Figure 2.7) captures a general CIS use-case with individuals assigned specific roles, and performed in a distributed environment with participants working on different (sub-) tasks on different machines in different locations.

Participants fulfill one or more roles with respect to both their task and organizational level, with such roles either explicitly assigned or implicitly inherited based on their expertise and experience [Prekop 2002] [S. Poltrock et al. 2003]. The model describes collaboration during work task performance along two dimensions: The work task dimension and the organizational dimension.

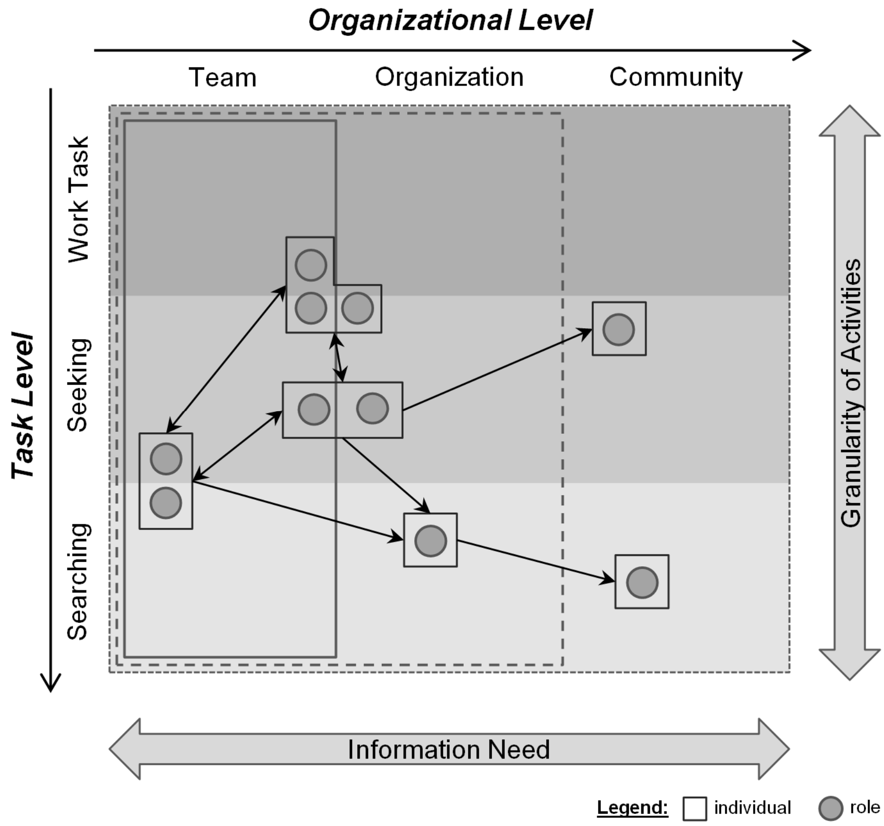



**Figure 2.7:** Model of Social and Work Tasks Aspects in CIS **[Böhm et al. 2014b]**

The granularity of collaboration activities changes with the task level. At the work task level, collaboration involves defining the problem and identifying the information need. At the IS level, it consists of formulating queries, assigning roles, and creating information paths [Prekop 2002]. At the information searching level, granularity becomes finer, for example, reusing previous searches, recommending documents, and sharing query terms or classification codes [Hansen and Järvelin 2005] [S. Poltrock et al. 2003]. The organizational level connotes participants' location in the collaborative network: the team that actually performs the task, the organization to which the team belongs, or a community outside the organization. The organizational level influences participant's behavior in terms of cultural norms, assigned responsibilities, knowledge, and other factors [Byström and Hansen 2005]. Information resources change in accordance with organizational level. Collaborators at the same level have access to the same information resources; if they need information outside of their organizational boundary, they must consult with or task others at that level to provide it [Prekop 2002].

## 2.4 Collaboration at Information Retrieval Level

The concept of CIR has been used in the past to refer to many different technologies that support collaboration in the IR process. Golovchinsky et al. provide a taxonomy that allows for structuring these different technologies [G. Golovchinsky et al. 2008]. This taxonomy incorporates the mode of collaboration as defined within the CSCW domain, i.e., describing collaboration in terms of the position in the two-dimensional space of time and location (see section 2.1.6) which is depicted in Figure 2.3. Moreover, the degree of coupling allows for further characterization of collaboration [Patel and Kalter 1993] [Haake and Wilson 1992]:

- Working without contact to any other team member is called ***Individual Work*** as part of a larger collaborative effort.
- Working with infrequent application state synchronization among team members is denoted ***Loosely Coupled*** mode.
- Team work situations in which team members share the same view on the same composite entity, i.e., synchronous application state, are called ***Tightly Coupled***.

Golovchinsky et al. propose additional dimensions for classifying CIR support systems. I will use the following definitions of these dimensions:

- Intent (explicit vs. implicit): ***Explicit Intent*** covers a team with members directly interacting with each other when resolving an information need. ***Implicit Intent*** is characterized by the provision of information based on similar information needs and based on other users' activities or opinions, without directly interacting with them.
- Information need (individual vs. shared): A ***Shared Information Need*** is characterized by a team with members having a declared understanding of the information on a topic that is required to achieve a team's goal. An ***Individual Information Need*** relates only to the goal of a single person.
- Depth of mediation (front-end vs. back-end): Depth of mediation is the level at which collaboration occurs in the system. A user interface may provide functions that allow for collaboration (***Front-end Mediation***), or the search engine

implements algorithms that incorporate team members' activities to modify the retrieval results (***Back-end Mediation***).

Based on the two dimensions Intent and Information Need, the following figure (Figure 2.8) summarizes sub-fields of research with respect to CIR. In the remainder of this section, these sub-fields are briefly introduced based on their position in this two dimensional space of Intent and Information Need.

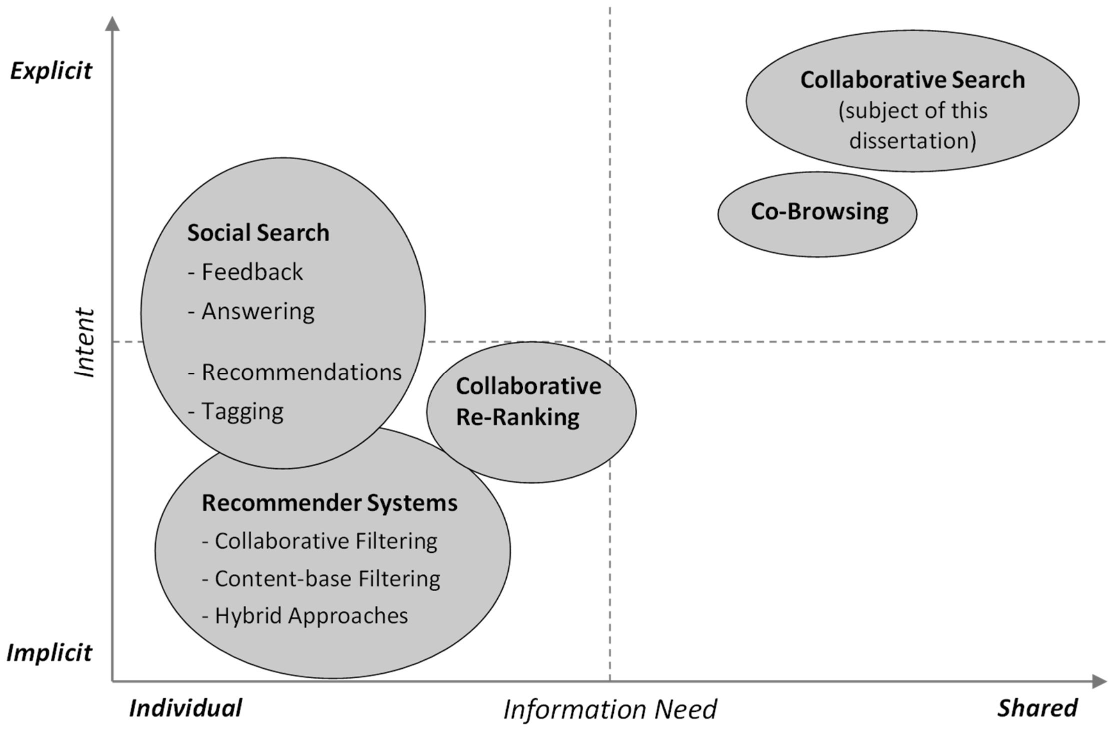



**Figure 2.8:** Sub-areas of research in the field of CIR; based on own perception.

### 2.4.1 Recommender Systems

***Recommender Systems*** have their origin in the field of ***Information Filtering*** [Hanani et al. 2001] which implement the idea of reducing the information overload [Eppler and Mengis 2004] by removing redundant or unwanted information from a set of items presented to the user. A recommender system is used to identify sets of items that are likely to be of interest to a certain user by exploiting a variety of information sources related to both, the user and the items. In contrast to information filtering, recommender systems actively predict in which items the user might be interested in and add those to the set of items provided to the user. Conversely, information filtering technology aims at removing items from such sets [Hanani et al. 2001]. Basis for such prediction are user profiles. Generally, ***User Profiles*** are the representations of users in an information system [Gauch et al. 2007]. Depending on the application, it may contain information about the users' vocabulary and interest. However, there are several ways to represent user profiles, such as weighted keyword profiles, semantic network profiles, concept profiles, etc. There also numerous ways to construct a user profile. A comprehensive review on this topic is given by [Gauch et al. 2007].

Recommender systems are usually classified into the following categories that reflect the way in which recommendations are made [Adomavicius and Tuzhilin 2005]:

- ***Collaborative Filtering*** algorithms build a user profile, based on activities (chosen items) or preferences (highly rated items). Items are recommended to people with similar profiles.
- ***Content-Based Recommendation*** approaches focus on building a representation of the content of the items in a system. Algorithms try to recommend items that are similar to those that a user preferred in the past.
- ***Hybrid Approaches*** combine collaborative filtering and content-based recommendation methods.

However, this is referred to as implicit collaboration, since although people may be generally aware that their results are based in part on data obtained from other users, but they may not know who those people were or what information need they had while searching. Thus, this kind of collaboration covers the re-use of historical data by the search engine as a source of evidence for document relevance [G. Golovchinsky et al. 2008].

### 2.4.2 Collaborative Re-Ranking

Collaborative approaches also attempt to discover patterns in the activities of a community of searchers in order to determine the general search context and prioritize search results accordingly. The I-SPY system [Barry Smyth et al. 2004] [B. Smyth et al. 2004] acts as a post-processing service for existing search engines that re-ranks results based on the learned preferences (e.g., histories of documents selected by users) of a community of users.

This is implicit collaboration, since the system creates personalization in an anonymous fashion. Even though community members are characterized by having a general common interest in a particular domain, users still do not know which community member contributes to the result and it's re-ranking. Also, other community members may have different information needs while searching.

### 2.4.3 Social Search

The general term ***Social Search*** has been applied to IR systems which utilize social cues provided by a large number of other people. Evans and Chi provided a general definition of social search:

> "*Social search is an umbrella term used to describe search acts that make use of social interactions with others. These interactions may be explicit or implicit, co-located or remote, synchronous or asynchronous*" [Evans and Chi 2008]

The following list provides examples of important services that aim to leverage a "*collective search experience*" [Evans and Chi 2008] and social network technologies during IR.

- ***Social Answering*** systems utilize people with expertise or opinions to answer particular questions in a domain [Chi 2008]. Answerers may come from various social contacts, such as friends, team members, and the greater public. *Yahoo! Answers*[1] is one example of such systems.
- ***Social Feedback*** systems utilize social attention data (e.g. votes of items of any kind) to rank search results [Chi 2008]. Feedback from users can be obtained either implicitly (e.g. obtained from user logs) or explicitly (e.g. ask users for votes, tags, and bookmarks).
- ***Social Tagging*** allows users to describe and categorize content for their own purposes using tags. Tags are keywords that describe characteristics of the item they are applied to and allow users to describe and organize content with any vocabulary they choose [Mathes 2004]. Users are free to apply any type and any number of tags to an object.

[1] To be found at: http://answers.yahoo.com

- ***Social Bookmarking*** websites allow users to store and access their bookmarks online through a Web-interface on a remote Web-server. The underlying application then makes all stored information shareable among users [Mathes 2004]. Additionally, such services are typically combined with social tagging.

Social search practices include explicit collaboration and implicit collaboration. However, they serve the purpose of satisfying the information need of a single person who is willing or able to involve a larger social network to satisfy that need. This larger social network may consist of various social contacts, such as friends, team members, or even the greater public. But either way, there is someone in that social network who already possesses the information that the initial individual requires. Therefore, the goal of a social search system is to correctly propagate existing information throughout the social network to create awareness for items that have already been discovered by at least one person [Gu and Han 2011].

### 2.4.4 Co-Browsing

***Collaborative Browsing*** (Co-Browsing) has been defined as "*... the activity of a team that surfs the Web with the goal of finding and retrieving information on a common topic*" [Cabri et al. 1999]. Co-Browsing systems extend conventional Web browsers to allow for joint navigation through the Web by two or more people accessing the same Web pages at the same time, i.e., work synchronously, and enabling real-time interactions among people working at the same time. Common to such systems is that collaborative Web browsing and real-time sharing of found information is supported by purposely built user interfaces rather than through external services like e-mail. For example, the CoFox system integrates communication and awareness services in the front-end [Rodriguez Perez et al. 2011].

In this way, Co-Browsing represents explicit collaboration with a shared information need supported by front-end mediated.

### 2.4.5 Collaborative Search

IR tasks in which two or more people are involved who lack the same information, i.e., share the same information need, and explicitly set out together to satisfy that need, are called ***Collaborative Search Tasks***.

Participants engaged in a collaborative search task aim to leverage each other's results as they explore a particular topic iteratively and interactively [J. Pickens et al. 2008]. A system supporting collaborative search provides mechanisms, interfaces, and mediation algorithms, that allow the team to work together to find information that neither individual would have found alone [Morris 2007]. There is an inherent division of labor and sharing of knowledge [Foley and Smeaton 2010] in collaborative work.

In the remainder of this dissertation, when referring to CIR, the specific characteristics of collaborative search (i.e. explicit collaboration and shared information need) are meant and thus, both terms, collaborative search and CIR, are used synonymously from here on.

## 2.5 CIR Support Systems

This section reviews research that was particularly concerned with the support of collaborative search tasks. In section 2.5.1, studies of experimental systems designed to support collaborative search settings are summarized. This is done in accordance to their degree of supporting communication, coordination, and cooperation between group members [Teufel 1995]. In section 2.5.2, I develop a schema for classifying software applications aiming to support collaborative information activities. Afterwards, research is reviewed and classified in accordance to the developed schema.

### 2.5.1 Studies of Experimental Systems and Settings

Computer-based collaboration environments typically provide a range of features supporting the basic requirements of coordination of team members and their activities, communication among team members, and collaboration on the development of shared artifacts.

#### 2.5.1.1 Communication

Communication support covers the provision of communication channels to team members, such as audio or video conferencing tools or text-based messaging systems. See Long and Baecker [Long and Baecker 1997] for a comprehensive taxonomy of computer-supported communication tools.

CIR support systems, such as Daffodil [Klas et al. 2008], SearchTogether [Morris and Horvitz 2007], Coagmento [Shah and Marchionini 2010], and CollabSearch [Yue et al. 2014] provide team members with integrated communication capabilities, such as instant messaging. Other collaborative environments employ external communication tools, such as Skype [Joho et al. 2008] [Villa et al. 2008]. However, those communication channels allow for discussion among team members about the work task, about information activities or any found item, and they allow for exchanging information objects, such as links, documents, queries, and search terms.

#### 2.5.1.2 Coordination and Awareness

Coordination support is usually realized using the provision of awareness mechanisms with the motivation being that when individuals are more aware of their team members' activities, they can coordinate the team activity themselves. ***Awareness*** is generally understood as the provision of information about activities of team members [Dourish and Bellotti 1992]. However, this implies two requirements. Firstly, awareness information has to make currently ongoing activities of interest visible to the users of a system. Secondly, it needs to provide an overview about changes in the past concerning the items of work. Fuchs et al. categorized types of awareness information along two dimensions [Fuchs et al. 1995]: (1) Concurrency, i.e., awareness regarding synchronous and asynchronous events, and (2) Task relation, i.e., awareness regarding task related and unrelated events.

Various awareness mechanisms have been implemented and evaluated in experimental CIR support systems. For example, WeSearch [Morris et al. 2010] and Fischlár-Diamondtouch [Smeaton et al. 2007] both supported team members in a synchronous, co-located search session at a table top. Both systems provided awareness information about query terms used and activities performed.

Evaluations of both systems revealed that the implemented awareness mechanisms did not only support coordination of team work, but also stimulated interactions between team members and discussions about search strategies among team members.

Other systems focused on improving the awareness across a distributed team of collaborating searchers, i.e., aiming at supporting remote located collaboration. To assist team members in completing their collaborative search tasks, awareness mechanisms provided information on the current and past information activities by providing query and visitation histories [Morris and Horvitz 2007] [Villa et al. 2008] [Shah and Marchionini 2010] [Yue et al. 2014]. Typically, users are also provided with a description of the current work task and are also provided with a shared workspace, where the team's saved information objects, such as documents, Web pages and snippets, are collected [Shah and Marchionini 2010] [Yue et al. 2014]. Additionally, notifications informed team members about recent events in the shared environment, such as recently added information objects to a shared workspace [Klas et al. 2008].

Evaluations of the latter systems revealed that missing awareness support led to fewer queries executed, fewer Web-pages examined, and an increased use of communication means [Shah and Marchionini 2010]. Whereas awareness has been identified as helpful when coordinating information activities and avoiding redundant work, awareness did not appear to help improve retrieval performance significantly [Villa et al. 2008].

#### 2.5.1.3 Collaboration

Collaboration support includes the provision of shared tools or application sharing functionality which enables the team members to either synchronously or asynchronously work on a shared artifact in order to achieve a shared team goal. Within CSCW research, shared artifacts are typically documents or working resources of any kind [Fuchs et al. 1995]. However, in the context of collaborative search, shared artifacts are typically the search product and the search process [Morris 2008] [Twidale et al. 1997].

CIR support systems provide an environment where collaboration between team members is mediated at different layers, also called depth of mediation [Golovchinsky et al. 2009], at which the mediation of the multi-user search process occurs.

Using front-end mediation [Golovchinsky et al. 2009], collaboration is supported via integrated functions in the user interface that focus on supporting explicit interaction between team members. Typically, those systems allow for exchanging queries and search results among team members through a shared user interface.

Using back-end mediation, each participants' activities are tracked and logged. Integrated algorithms in the search engine evaluate these activities and combine them algorithmically to produce retrieval effects that follow some defined strategy. The intention is to allow team members to work independently but still be influenced by their team mates by incorporating their activities in the result-sets (merging, splitting) and queries (term re-weighting).

Joho et al. as well as Foley and Smeaton explored the potential benefit of adopting several IR techniques, such as search result division and relevance feedback, to support division of labor and sharing of knowledge among team members [Joho et al. 2009] [Foley and Smeaton 2010]. For example, the search engine implemented a division of labor strategy by preventing documents to be provided to users as query response that were inspected by a team member already. Sharing of knowledge was realized by re-weighting of query terms that occurred in documents judged as relevant by team mates.

Other approaches explored regrouping of search results based on user roles that were manually predefined. In [J. Pickens et al. 2008], back-end mediation was implemented based on team members fulfilling asymmetric roles (Prospector and Miner) in a synchronous search session. The search engine performed a re-ranking of a result list based on judgments of all participants. This re-ranking was based on the two measures relevance (the ratio of relevant and non-relevant documents in a response list) and freshness (the ratio of inspected to non-inspected documents) in a result list. Documents not inspected by the Prospector were forwarded for examination by the Miner. Similarly, Shah et al. [C. Shah et al. 2010] also investigated merging and splitting of query results among team members with asymmetric roles (Gatherer and Surveyor). However, these studies restricted user roles into predefined categories. To provide more flexible CIR support, Soulier et al. aimed at mining such user roles in a collaborative search session to leverage diverse sets of knowledge present in the team [Soulier et al. 2014].

Common to all these examples is that the search system realized an information flow between the participants that do not have to manually decide how to divide the IR task and which documents to inspect.

Evaluations of front-end mediation approaches showed mixed results. For example, Mitrelis, Tsakonas, and Papatheodorou did a user evaluation of Daffodil [Klas et al. 2008] and found that making annotations, comments and recommendation helped in the advancement of the collaborative search task of the team members [Mitrelis et al.

2008]. Also, Morris and Horvitz identified that being able to make recommendations to team members was highly rated by users [Morris and Horvitz 2007]. However, automatic division of labor features, such as *"split search"* and "*multi-engine search"*, were not heavily used by study participants and opinions were divided on the utility of these features [Morris and Horvitz 2007]. Additionally, user feedback indicated that collaboration means integrated in the front-end led to difficulties in usage of the system's functionality due to the complex user interface [Mitrelis et al. 2008].

Back-end mediation of collaboration in IR has been investigated in [J. Pickens et al. 2008], and [C. Shah et al. 2010] as well as in [Foley and Smeaton 2010], [Joho et al. 2009], and [Soulier et al. 2014]. Experiments showed that back-end mediation led to an improvement of the collaborative performance. This covered retrieval performance, i.e., the system allowed a team to find relevant information more efficiently and effectively, as well as exploration, i.e., the system allowed a team to find relevant information that cannot be found while working alone [J. Pickens et al. 2008].

### 2.5.2 Collaborative Information Activities

Foster presented a literature review describing research related to collaboration during IS tasks and IR tasks [Foster 2006]. Foster identified that CIR technologies support specific stages of the IR process, such as query construction: However, Foster also analyzed collaborative filtering in his survey. So, Foster did not distinguish between explicit collaboration and implicit collaboration and also did not cover all stages of the search process.

This sub-section presents a more detailed analysis and classification of recent work in the area of collaborative information activities. For the purpose of this dissertation, I define ***Collaborative Information Activities*** (CIA) as information activities performed by team members that share the same information need. As basis for this classification, I used the model developed by Landwich, Klas et al. to describe an individual's information activities. In [Landwich, Klas, et al. 2009], the authors pursued the approach of an interactive information dialogue cycle as developed in [Hemmje et al. 1996]. They describe the information searching process as a dialogue between user and system consisting of six activities that were assigned to three stages (the so called interaction modes of the user):

1. ***Access***: Query construction and submission (Exploration),
2. ***Orientation***: Move within and refinement of the result set, change of focus (Focus, Navigation, Inspection),
3. ***Assessment***: Identification of relevant information objects (Evaluation, Store).

The dialogue cycle starts with a first query and ends after *n* cycles with a resolved or at least reduced information deficit.

#### 2.5.2.1 Access

During Access, users are able to benefit from their team members by exchanging query definitions and modifying and executing them for their own purposes. This is realized in different ways.

With ***Query Re-Use*** I refer to the activities that realize the exchange of (full) query definitions between team members. The team members are able to perform the exchange interactively by

- Choosing the query definition from a shared repository [Romano Jr et al. 1999; Walkerdine and Rodden 2001],
- Choosing the query definition from the query-history of another team member [Morris and Horvitz 2007] [Shah and Marchionini 2010], or
- Exchanging the query definition as separate, persistently stored object [Klas et al. 2008] [Twidale et al. 1997].

With ***Group Feedback*** I refer to a group-based adoption of relevance feedback methods. This class of collaborative activities incorporates the explicitly or implicitly provided relevance judgments of the team members. As result, queries are modified accordingly by adapting the weights of the query terms or expanding the query with additional query terms. This includes various approaches of query expansion techniques that generally extract search terms from highly ranked documents of previously issued queries.

Query terms might be estimated algorithmically in different ways:

1. The systems extract terms from highly ranked documents of previously issued queries that are similar to the current query [Fitzpatrick and Dent 1997; Hust 2005].
2. Query terms are extracted from queries (generated by other users) that are associated with documents of the current result set [Billerbeck et al. 2003].

#### 2.5.2.2 Orientation

During Orientation, division of labor strategies are implemented using ***Result-Set Splitting***, i.e., algorithmic division of a search result among the team members. The result set of a query is distributed algorithmically among the team members. These sub-sets are typically disjoint, i.e., the participants will only obtain documents that no other member has seen before [Foley and Smeaton 2010]. This splitting of search result sets can further be based on specific roles that are assigned to the participants, such as Prospector and Miner [J. Pickens et al. 2008], or based on personal relevance, i.e., thematic focus and interests of the participant [Morris et al. 2008].

In addition to this, result sets can be enhanced algorithmically or manually using documents identified by other team members. ***Result-Set Merging*** is based on the similarity between a user profiles and queries: Documents returned by previous queries and judged as relevant by team members will be added to the result set of a recently executed query [Naderi et al. 2007]. ***Document Recommendation*** includes the interactive recommendation of documents or links. Documents that have been identified by other participants and estimated as possibly interesting for another team member, are recommended and added to the work list of a team member [Klas et al. 2008] [Shah and Marchionini 2010].

#### 2.5.2.3 Assessment

During Assessment, collaboration addresses the diversity of domain knowledge across the group: ***Combination of Judgments*** refers to the combination of the different document assessments of the group members. The relevance of a document is determined by the opinions of multiple users through interactive voting: In [Romano Jr et al. 1999] a scale-based approach is implemented, in [Capra et al. 2012] a traffic light based approach is used. ***Re-Ranking*** refers to the algorithmic re-ordering of the results. The ranks of the search results are determined not only by the relevance to the individual user, but also by the relevance to the entire group. This might be realized by using term frequencies in the stored objects or bookmarks of team members [Morris et al. 2008].

#### 2.5.2.4 Summary

Research aiming at supporting collaboration at IR level has conceptualized, implemented, and evaluated collaborative applications and services for use at each stage of an IR process. Figure 2.9 depicts an integrated model of the Information Dialog developed by Landwich, Klas et al. and the task model developed in [Byström and Hansen 2005]. Figure 2.9 depicts the classes of activities available for a team member to collaborate with the rest of the team at the IR level of the work task.

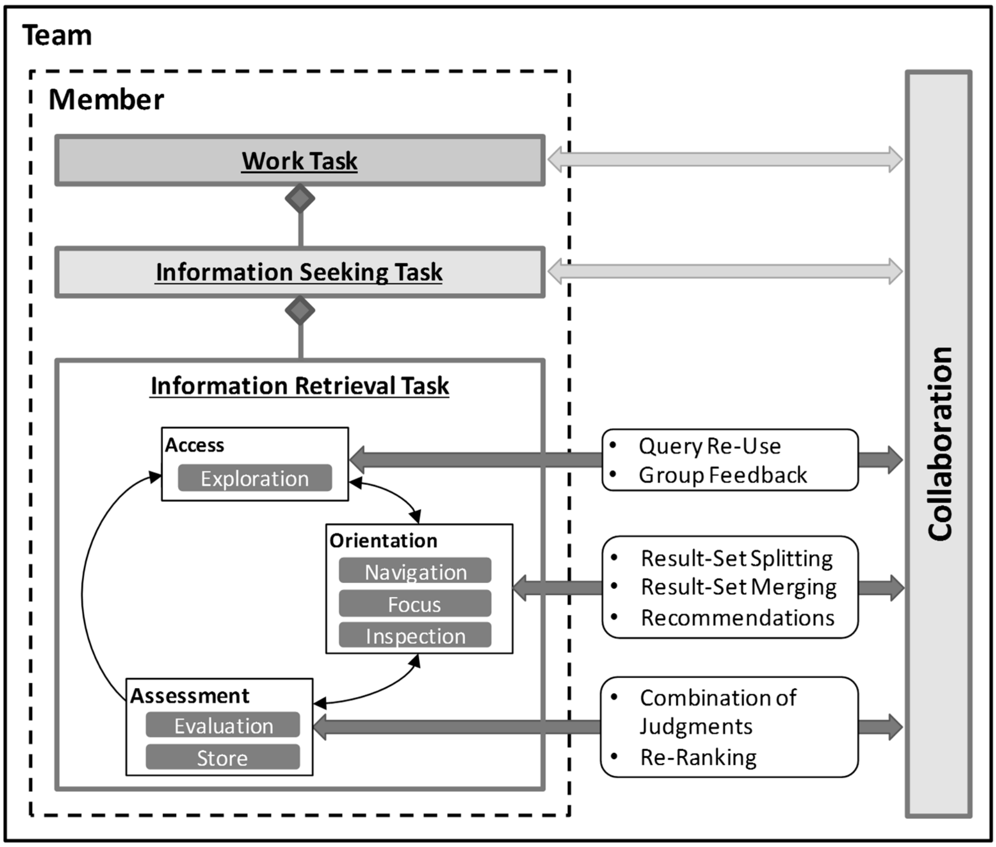



**Figure 2.9:** Summary of collaborative information activities **[Böhm et al. 2013]**

## 2.6 Evaluation of CIR

This section presents several approaches towards evaluation of IR as well as CIR systems. I introduce the underlying concept of relevance in section 2.6.1. Moreover, problems with IR evaluation and the methodologies used in standard IR evaluation are discussed (see section 2.6.2). Afterwards, methodologies adopted towards CIR evaluation are presented (see section 2.6.3) and, finally, measures used for CIR evaluation are defined in section 2.6.4.

### 2.6.1 The Notion of Relevance

The final stage in the development of any IR system is its evaluation. The fundamental concept in IR is that of relevance which is regarded to as the extent to which a document answers a stated information need. However, the concept of relevance is not clearly defined in IR. For example, Saracevic conducted three reviews (in 1975, 1996 and 2007) of the progress in thinking about the nature of relevance in Information Science [Saracevic 1975; Saracevic 2007; Saracevic 1996]. Mizarro [Mizzaro 1997] discussed about 160 papers in an attempt to define various aspects of relevance.

This section briefly reviews the most popular arguments about the nature of relevance. Exhaustive reviews have been conducted by Saracevic [Saracevic 1975; Saracevic 2007; Saracevic 1996] and Mizarro [Mizzaro 1997].

According to Saracevic, the fundamental idea of IR was and still is to retrieve relevant information with the help of technology. Thus, relevance became the central notion of research which Saracevic classified as follows:

- ***System Relevance*** (also known as algorithmic relevance): the relationship between a query and the objects (typically documents) retrieved by the system.
- ***Topical Relevance*** (also known as subject relevance or topicality): the relationship between the topic expressed by a query, and documents that are about that topic.
- ***Pertinence*** (also known as cognitive relevance): the relationship between a user's information need, taking into account the user’s current background knowledge, and information addressing that information need in the system.
- ***Utility*** (also known as situational relevance): the relation between a situation at hand and information in the system. This differs from pertinence in that it covers more than just the specific information need, and takes into account, e.g., the extent to which the user can make use of the information and the extent to which the information reduces uncertainness regarding the situation.

- ***Motivational Relevance*** (also known as affective relevance): the relationship between the users' ultimate goal, intents and motivation and the information in the system. If the user is satisfied, the goal is accomplished etc., then the information is motivationally relevant.

However, relevance is an inherently subjective notion and users may differ on their assessment of the relevance or non-relevance of documents w.r.t. the same information need. Saracevic's survey of the literature reveals four major findings on relevance judgments:

1. Users make relevance judgments based on different document attributes (content is one of them)
2. The attributes that matter most depend upon the user's internal and external context
3. Context varies across users, so relevance judgments vary across users
4. Context varies over time, so relevance judgments (for the same user) vary over time

Judgments are made within a context, i.e., the internal context (users' knowledge, feelings, and expectations about the information need, the corpus, and the system), and the external context (users' higher-level task at hand and their environment). These contexts are dynamic, so "*relevance is dynamic across users and for the same user across time*" [Saracevic 2007].

### 2.6.2 Information Retrieval Evaluation

Assessing the quality of IR systems is the process of measuring the effectiveness, i.e. assessing how well a system meets the information needs of its users [Sanderson 2010]. However, IR evaluation is twofold, user-oriented and systems-oriented. The former one focuses on evaluating how effectively and efficiently users' search for information meets their needs, helps solving a problem, accomplishing a work task, or achieving the users' overall goals [Cole et al. 2009]. From the systems perspective, IR evaluation is focused on evaluating the effectiveness and efficiency of the retrieval system, e.g., for a query and a list of returned documents one has to check whether the returned documents are relevant [Sanderson 2010].

System-oriented evaluation, such as embodied by the Cranfield and TREC [Sanderson 2010] evaluation models, makes a number of simplifying assumptions about users, their needs, and behaviors. It holds the number of model-variables small to be able to analyze the subject of interest as precisely as possible. It provides means for quantifiable evaluation and comparison of system components [Sanderson 2010]. It evaluates IR systems with a static and reusable test collection which consists of a document corpus, a set of statements of information needs (called topics) against the document corpus, and manually annotated relevance assessments indicating the relevance relations between the topics and the documents. The evaluation process requires a measure that is applied to compare a query-response (retrieved documents list) and the relevance assessments for each topic [Sanderson 2010]. The measured score is most often used as an indicator of the performance of one system relative to another, with the assumption that similar relative performance will be observed on other test collections and in operational settings [Sanderson and Zobel 2005].

However, the test-collection based evaluation paradigm has often been criticized for being of limited value for assessing IR applications, as it does not account for presentation issues, intellectual and physical user effort [Hansen and Järvelin 2005]. Also, there is evidence that optimizing IR engines using batch experiments in the ad-hoc style of TREC does not necessarily translate into an improved IR application for users [Hersh et al. 2000]. However, test-collection based evaluation does provide adequate means for comparing the relative effectiveness of two retrieval strategies (or system components, or algorithms) [Voorhees 2000]. As the latter is of interest in this dissertation, system-oriented evaluation will be analyzed in more detail in this section.

### 2.6.3 Evaluation Methodologies for Collaborative Search

Evaluating the performance of CIR support systems is more challenging than the evaluation of IR systems design towards individual use. This results from the complex and dynamic interactions that take place among team members and system [Shah 2014]. Rather than having one user and its information activities, at any point in a collaborative search session, there are several users to consider and each of them may have submitted queries, obtained results, and assessed documents.

Baeza-Yates and Pino first presented initial work on measures that extended the evaluation of a single-user IR system for a collaborative environment [Baeza-Yates and Pino 1997]. Their approach was based on dividing a collaborative search session into several stages. Team member's search results were accumulated after each search stage. Baeza-Yates and Pino extended traditional retrieval measures recall and precision by treating the performance of a group as the summation of the performances of the individuals in the group.

Shah [Shah 2014] proposes that evaluating various factors in CIR behaviors and results can be summarized as measuring the retrieval performance of the system, the effectiveness of the interface in facilitating collaboration, and the user satisfaction and involvement.

This section summarizes methodologies that have been used to evaluate CIR support systems. In the following sub-section, these methodologies are clustered according to [Shah 2014], that is, into three major approaches (1) User Studies, (2) System-based Testing, and (3) Ethnographic Field Studies.

#### 2.6.3.1 User Studies

The effectiveness of a CIR system has often been evaluated by looking at the usability of the collaborative interface in a laboratory setting, where subjects were often asked to perform a certain task and were provided with a search system and a set of collaborative tools [Crescenzi and Capra 2013] [Shah and Marchionini 2010] [Shah and González-Ibáñez 2010] [Shah and González-Ibáñez 2011] [Shah and González-Ibáñez 2012] [Morris and Horvitz 2007] [Mitrelis et al. 2008]. Data for evaluation has been gathered via the logs of the information activities, observations and questionnaires. Shah [Shah 2014] summarizes typical elements of study design and evaluation methodology for those approaches as follows:

- A controlled setup like a lab,
- Selective group of participants,
- Supervised or semi-supervised execution of task,
- Recorded data, such as logs, questionnaires, interviews

Partly, study methodologies are brought from the field HCI, and analysis involves quantitative as well as qualitative approaches. It is very common that the aim of user studies was to measure aspects of usability [Shah 2014] [Nielsen 1994]. Sometimes studies have been conducted to simply see how the users like a new CIR interface. Other times, the research interest covered questions about how users utilize various tools offered in their interface and how those tools affect the act of collaboration and how it impacts their work [Shah 2014].

#### 2.6.3.2 System-based Testing

CIR support systems were often assessed by employing various system-oriented measures for evaluating the potential retrieval effectiveness resulting from the use of that system. Those approaches often used simulations of user interactions and/or batch processing [Soulier et al. 2013] [C. Shah et al. 2010] [Barry Smyth et al. 2004] [Foley and Smeaton 2010] [J. Pickens et al. 2008]. For example, Pickens et al. [J. Pickens et al. 2008] showed how their algorithm could achieve an effective collaboration by way of simulation, Shah et al. [C. Shah et al. 2010] demonstrated how search processes that were virtually combined could result in achieving results that are both relevant and diverse. Foley and Smeaton [Foley and Smeaton 2010] as well as Soulier et al. [Soulier et al. 2013] demonstrated the effectiveness of their models by simulating users searching together synchronously based on interaction logs of individual users from the TREC interactive track experiments.

According to Shah [Shah 2014], typical elements of study design and evaluation methodology for system-based experiments involve:

- A standardized test-collection consisting of a document corpus, topics/queries as well as relevance judgments,
- An approach that typically involves an algorithm/system that provides system-mediated collaboration.

Experiments were run using partial or full simulations and analysis typically involved quantitative approaches. Simulations of users' interactions have been found suitable in CIR research in cases where CIR support systems with back-end mediated collaboration were involved.

#### 2.6.3.3 Ethnographic Field Studies

Ethnographic approaches usually focused on behavioral aspects with qualitative analysis. Such studies have been conducted in several domains of work, such as engineering, IP domain, medical care, academic research and higher education

According to Shah [Shah 2014], typical elements of study design and evaluation methodology for ethnographic/field studies involve:

- Data collection through observations, surveys, and interviews,
- Semi-supervised to unsupervised execution of task.

Such studies usually last several days to several months. Typical study instruments are brought from social sciences and analysis typically involves qualitative approaches. Hence, such evaluation approaches are not appropriate means for comparing retrieval functions or algorithms in a system. However, typical research question addressed by such studies is: *"How is group member behavior related to contextual factors (work task)?"* [Hyldegard 2006]

#### 2.6.3.4 Discussion

The summarized evaluation approaches show that evaluation with respect to CIR support systems is also twofold and depending on the research focus. System-oriented evaluations: With the aim of enhancing the productivity in an information searching process, measuring the outcome of the process is often performed by measuring the increase of retrieved relevant results [J. Pickens et al. 2008]. This is a popular approach for research that focuses on algorithmic or back-end mediation [G. Golovchinsky et al. 2008]. In contrast, user-oriented evaluations employ instruments taken from HCI and concentrate on the usability [Nielsen 1994] of CIR environments where the typical questionnaires relating to user interactions are analyzed in the context of the usage data collected through logging.

### 2.6.4 Measures

As proposed in [Baeza-Yates and Pino 1997], measures used in evaluation of CIR systems are often based on extended traditional retrieval performance measures, namely recall and precision (@N) [C. Shah et al. 2010] [Barry Smyth et al. 2004] [Foley and Smeaton 2010] [J. Pickens et al. 2008]. Shah and Gonzáles-Ibánez [Shah and González-

Ibáñez 2010] and Shah [Shah 2014] proposed the following enhanced measures for evaluating the effectiveness of a CIR support system, classified into system-oriented and user-oriented measures.

#### 2.6.4.1 System-oriented Measures

Traditional IR evaluation metrics [Sanderson 2010] are precision, recall, and F-measure. These measures have been adapted by considering a team $T$ of $N$ team members that participating in a collaborative search session. Shah and Gonzáles-Ibánez define the *Coverage(T)* as the union of documents visited by any team, and similarly, *RelevantCoverage(T)* as the union of document that a team visited and found as relevant. Based on these definitions, traditional measures are redefined as follows:

$$Precision(T) = \frac{RelevantCoverage(T)}{Coverage(T)} \quad (2.1)$$

$$Recall(T) = \frac{RelevantCoverage(T)}{U_r} \quad (2.2)$$

$$F = \frac{2\, Precision(T)\, Recall(T)}{\left(Precision(T) + Recall(T)\right)} \quad (2.3)$$

Here, $U_r$ denotes the universe of relevant documents. Given a test collection, $U_r$ corresponds to the documents annotated relevant given a specific topic. However, for Web-search tasks, as discussed in [Shah and González-Ibáñez 2010], $U_r$ is determined by the union of relevant documents found by all teams involved in an evaluation, i.e., this assumes that the user study is conducted involving several teams (see [Shah and González-Ibáñez 2010] for an example).

Furthermore, to allow analyzing the contribution of each team member, Shah and Gonzáles-Ibánez propose measuring coverage of information by each team member. For this purpose, *IndividualCoverage(τ)* and *IndividualRelevantCoverage(τ)* are introduced and describe the disjoint (relevant) sets of documents discovered by one team member $\tau \in T$.

#### 2.6.4.2 User-oriented Measures

Shah [Shah 2014] proposed measuring various aspects of communication in a collaborative search session. In [Shah and González-Ibáñez 2010], a coding schema for classifying exchanged messages between co-workers is proposed as well as corresponding measures. The coding schema consists of a group of four major categories of messages: (1) Task coordination, (2) Task content, (3) Task social, and (4) Non-task related. In addition to classifying messages according to the coding scheme above, the authors described a set of quantitative measures that can be used to measure the balance and effort of the interactions during the collaboration process, i.e. communication volume (the overall number of messages issued by an individual team member during the collaboration process), and the communication effort (proportion between the number of words in each message and average words per minutes).

Additionally, to perform further user-oriented evaluation, questionnaires were proposed to gather qualitative user responses [Shah 2014].

## 2.7 Discussion and Identification of Remaining Challenges

IR has traditionally been considered as solitary activities: A single user identifies an information need, uses an IR system to discover relevant information, and iteratively resolves the information need. Numerous studies have found, however, that such activities often involve multiple people. These studies have been conducted in engineering [Bruce et al. 2003; Morten Hertzum and Pejtersen 2000; Steven Poltrock et al. 2003], legal practice [Attfield et al. 2010], IP domain [Hansen and Järvelin 2005], medical care [Reddy and Jansen 2008], academic research [Spence et al. 2005], higher education [Hyldegard 2006; Talja 2002; Twidale et al. 1997], and military [Prekop 2002]. Such observational works focused on behavioral aspects and are valuable at helping to describe how users behave and interact in various situations and conditions. These works have identified broad patterns of IIB and CIB. However, they did not provide guidance to design systems that support collaboration in IR.

Effective and efficient collaborative performance of IR tasks in distributed environments requires combination of expertise, special skills and knowledge of team members to allow for efficient achievement of goals [Sonnenwald et al. 2004]. A major challenge for a CIR support system is to provide awareness on the best suited collaboration partners at a given stage of work task performance. Fidel et al. [Fidel et al. 2000] recommend that systems designed to support CIR should incorporate features that support interactions between users and "*enhance their access to one another's knowledge, ideas, and opinions or help them keep on track*" ([Fidel et al. 2000], p. 951).

Previous research on CIR support systems has focused on algorithmic support of collaboration (back-end mediation) as well as on improving human-human and human-computer interaction by facilitating communication, coordination and awareness mechanisms (front-end mediation). These experimental systems realized a tightly coupled collaboration, i.e., each participant knowing continuously about activities made by others.

CIR support systems typically provided a team with a shared workspace and usually also provided special tools for supporting communication, coordination and cooperation [Teufel 1995]. Conducted studies investigated the influence of live communication channels [Klas et al. 2008] [Morris and Horvitz 2007] [Shah and Marchionini 2010] [Joho et al. 2008] [Villa et al. 2008] and awareness information on team work [Morris et al. 2010] [Smeaton et al. 2007] [Klas et al. 2008] [Morris and Horvitz 2007] [Shah and Marchionini 2010] [Villa et al. 2008] as well as the cognitive load resulting from the additional awareness information presented to users [Shah and Marchionini 2010] [Shah and González-Ibáñez 2010]. Such approaches have been phrased front-end mediated collaboration [G. Golovchinsky et al. 2008]. Common to these approaches is that searchers may collaborate at the user interface and interpersonal level, however, the search engine itself does not support collaboration. Instead, searchers are automatically notified about the on-going activities of their team mates, but to take advantage of that information and to improve their searches, each user must manually examine an interpret team mates' queries and retrieved documents (or visited Web-pages, respectively).

While awareness has been recognized as important feature towards collaborative search, solutions based on the user interface only require the attention to others' information activities, i.e., searchers must manually reconcile and integrate their activities with those of their team mates [J. Pickens et al. 2008]. Conversely, using algorithmically mediated collaborative search [G. Golovchinsky et al. 2008], the system's back-end coordinates user activities throughout the session. This has been phrased back-end mediation. Evaluations suggests that an algorithmic mediation approach allows teams finding more unique relevant documents, i.e., increases retrieval performance of a team [J. Pickens et al. 2008]. However, back-end mediated collaborative search has only been studied for synchronous and co-located settings and specific search tasks, such as video retrieval [Gene Golovchinsky et al. 2008] [Adcock and Pickens 2008].

Recent empirical studies [Morris 2013] [Kelly and Payne 2014] revealed that, despite the increasing availability of systems that are specifically designed to support CIR, in professional practice, search systems and interfaces designed for individual usage are utilized in collaborative work [Morris 2013] [Kelly and Payne 2014]. In the collaboration environments of professional practice, missing awareness support of team member's activities calls for alternative approaches for effective coordination of team work. Studies of CIR in professional settings showed that differences in knowledge and skills affect the way in which knowledge workers organize and perform work [Fidel et al. 2000]. This highlights the need for services that aim at a more flexible and adjustable activity coordination among team members that account for differences in skills and knowledge.

Collaborative search in professional settings is characterized by team members working independently, i.e. in a loosely coupled mode, and utilizing a heterogeneous work environment, and working remote located and potentially asynchronous. For such settings, it is an open problem how existing user support principles could be enhanced to support team work. Typically, IR models consider individual users when, e.g., optimizing and ranking search results. Generally, IR models for collaborative search are an underexplored research area. Thus, in this dissertation, I develop informal as well as formal IR models towards support of a multi-user, collaborative search session to address the remaining challenges.

# 3 Modelling CIR Support

As it has been outlined in chapter 2, there are many different CIR settings each of which characterized by specific conditions and requirements so that it is unlikely that there is a *"one-fits-all"* system solution. To be able to adequately develop models for CIR support, the corresponding setting and technical prerequisites that are in the focus of this dissertation need to be stated clearly. To that end, a user study has been conducted that aimed at capturing the technical environment in which collaborative IR activities are performed. The findings of this study are incorporated into informal models that provide a structured outline of the entities involved in a CIR task. Moreover, these informal models help at tailoring the formal model for CIR support eventually developed in this chapter.

The remainder of this chapter is structured as follows:

Section 3.1 reports on a pilot user study that investigated the CIR practices of three work groups in academic and industrial research facilities. A further analysis of the results and the derived informal models are presented in section 3.2. These informal models cover the conceptual system models of CIR environments as well as the process of CIR support to be provided in such environments.

In section 3.3, a formal cost-model for collaborative search is developed that serves as basis for deriving a novel ranking principle for collaborative search sessions which is also presented in this section.

Finally, section 3.4 introduces the notion of collaborative ***Activity Suggestions***, that is, a formal criterion that is also derived from the cost-model developed in section 3.3. This formal criterion describes optimum collaboration strategies in IR as the solution of an integer linear program. It accounts for different information activities performed within the team and allocates documents to team members accordingly, i.e., estimates which document should be inspected by whom.

This chapter closes with a discussion of the developed models and the limitations (see section 3.5). This discussion includes a brief statement on the estimation of the models' parameters as well as the models' relation to previous CIR approaches and the PRP.

## 3.1 Study of CIR Practices

To approach research question *RQ1* (see chapter 1), a user study has been chosen as appropriate method to obtain concrete numbers that express preferences, trends, and demographics. Whereas the studies and systems discussed in chapter 2 commonly focused on information needs as well as IR processes within teams, less attention has been devoted to how people utilized current technology to realize CIR. The conducted user study did not aim at analyzing the CIR activities in detail but rather it aimed at capturing the use of software technologies for realizing collaboration, IR, and information sharing in real-world settings.

Similar to the online survey conducted by Crescenzi and Capra [Crescenzi and Capra 2013], I made implicit assumptions about the components involved in the collaborative processes. Those were: (1) A ***Search Component*** in which team members conduct searches to retrieve information, (2) a ***Communication Component*** in which team members coordinate their activities and communicate regarding the search process, and (3) an ***Information Sharing Component*** in which collaborators share their search products.

This section starts with detailing the study method (see section 3.1.1). Afterwards, the results are presented (section 3.1.2) and discussed (section 3.1.3). This survey and its results have also been published in [Böhm et al. 2013] and [Böhm et al. 2014a].

### 3.1.1 Survey Method

Nowadays, scientists have a wide variety of software applications at their disposal to meet the daily work demands. To identify which technologies and means that constitute the collaborative environment used by researchers to perform collaborative work tasks, an online survey (implemented using Google Drive) has been conducted. I invited researchers to answer questions regarding the acquisition of required information with respect to the collaborative performance of their work tasks. In addition to questions regarding demographics, I was particularly interested in how they (1) collaborated with colleagues when performing an IR task, (2) communicated with their colleagues and shared information, and (3) how they identified colleagues who could be most helpful in regard to answering their questions and solving problems.

In November 2012, I asked members of two work groups of a university research facility (each in the field of life sciences) to complete my survey, and in May 2013 I asked the members of an industrial research department (in the field of IT). The survey has been provided via e-mail distribution lists addressing (in sum) 52 people. 24 completed the entire survey, yielding a 46.2% response rate. The survey consisted of both free-text and multiple-choice questions.

### 3.1.2 Survey Results

#### 3.1.2.1 Demographics

The age of the participants ranged from 24 to 43, with an average age of 33.25 years (standard deviation 5.14). 75% of respondents were male. Respondents were specialized in different fields of study. I wanted to estimate the degree of experience the respondents had in collaborating with colleagues. The number of articles published by multiple authors is often seen as a measure of research collaboration [Bukvova 2010]. I decided to use this measure although not every research collaboration results in a publication and not all co-authored articles are result of collaborative research [Bukvova 2010]. Participants were asked for the number of co-authored writings (papers of all types, grant application, project reports, etc.) they had contributed to. The participant's responses covered a broad range of quantities and thus yielding a large standard deviation (s.d.) of 23.8. The average number of co-authored writings was 18.73.

Additionally, I asked for the highest academic degree: 9% of the respondents hold a Bachelor's Degree (or equivalent), 26% of the respondents hold a Master's Degree (or equivalent), and 61% of the respondents hold a Doctor's Degree (or equivalent). The remaining 4% were Students before their first academic degree. Participants were asked to self-rate their search experience. On a five-point Likert scale, 4% rated themselves as inexperienced, 13% as moderately experienced, 67% as experienced, and 13% as expert. No respondent self-rated as *"very inexperienced"* user. Results showed that, in addition to the high level of familiarity in search practices, the group of respondents was characterized by high degree of education, research, and collaboration experience.

#### 3.1.2.2 Search Habits and Result Management

Participants were asked about the (electronic) information sources they frequently used (see Figure 3.1) as well as software applications utilized to organize and manage their search results (see Figure 3.2), i.e., scientific literature. Respondents could select electronic information sources in a multiple choice box. Additionally, they were able to extend this list by naming further software applications (*"Other"*).

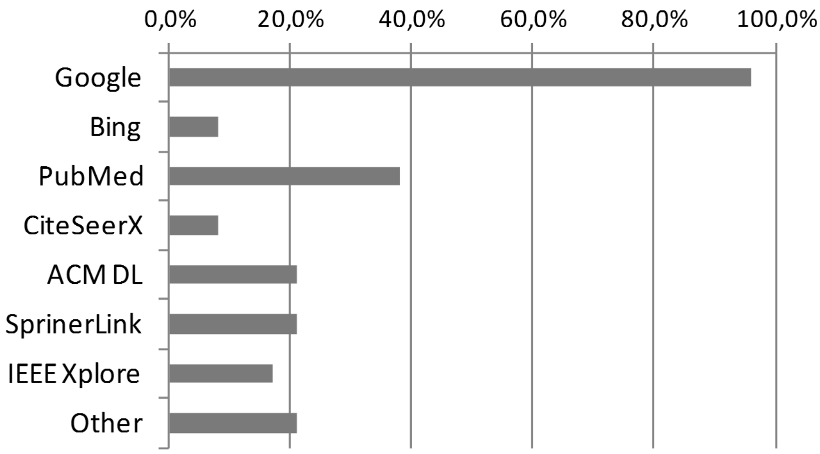



**Figure 3.1:** Electronic information sources used by respondents

Figure 3 summarizes the selected sources of information. *"Other"* included *Microsoft Academic Search*, *Zentralblatt MATH*, *DBLP*, and *Ecosia*. The results show Google as a common favorite choice, but they also provide evidence of the diversity of electronic information sources consulted by respondents. Figure 4 summarizes the selected literature management software. Others were: *www.citemaster.net*, *BibTeX*, *Citavi*, and the *Windows Explorer*. In total 10 distinct software applications have been named by the respondents. This too points to a broad variety of software tools in operation.

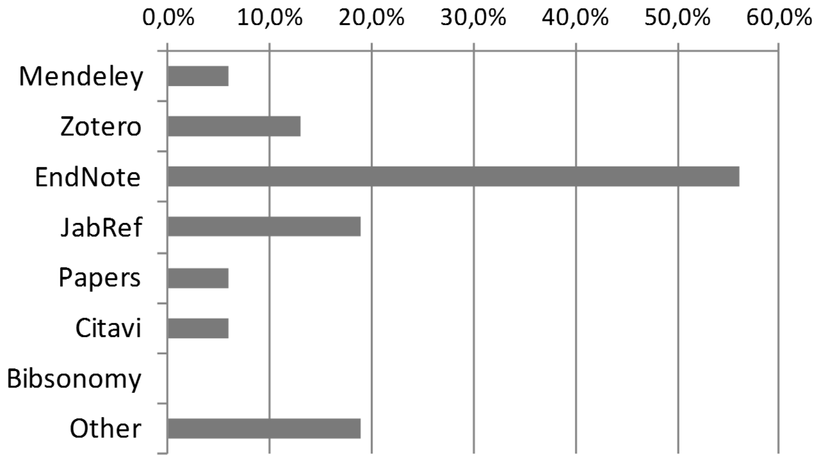



**Figure 3.2:** Literature management applications used by respondents

3.1.2.3 Collaboration during Search

To learn more about practices of collaboration during search, I asked the participants in which stages of the search process they consulted colleagues or have been consulted. I asked regarding collaboration during information source selection (Q1) and query formulation (Q2a and Q2b). According to Marchionini, query formulation involves (a) an action mapping of the information searcher's strategies and tactics onto the features the system interface provides, and (b) a semantic mapping of the information searcher's vocabulary onto the system's vocabulary [Marchionini 1997]. Therefore, I included questions addressing collaboration with respect to the interface and its functions (action mapping, Q2a) as well as collaboration regarding the query formulation (semantic mapping, Q2b). Furthermore, Q3 and Q4 addressed the search result refinement as well as the search result assessment. Figure 3.3 depicts the questions and respondents' answers.

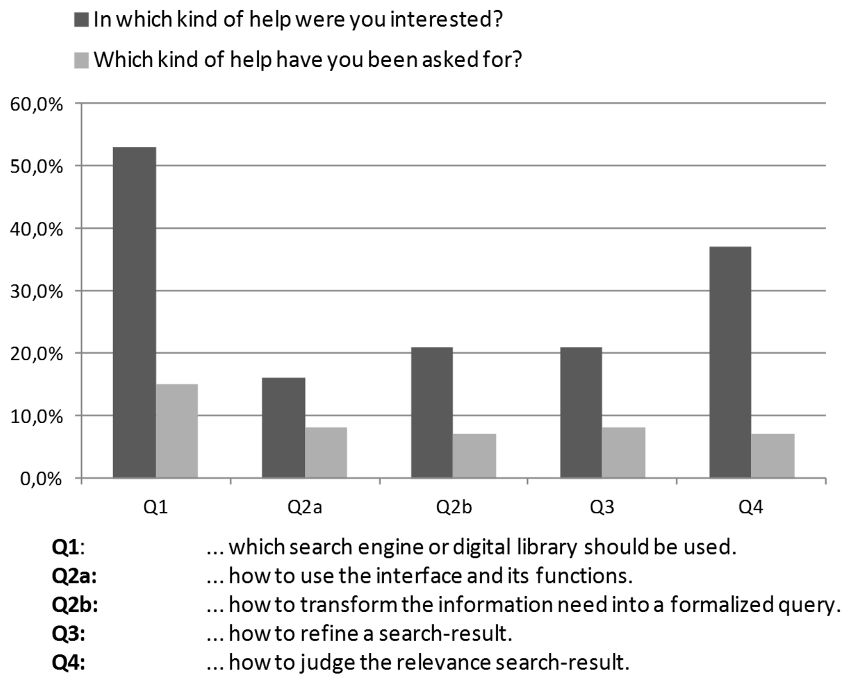



**Figure 3.3:** Percentages of respondents who collaborated during different search stages

Collaboration was found to be at its highest during the information source selection stage as well as during the assessment stage. However, generally, collaboration can be identified in each stage of the search process.

#### 3.1.2.4 Communication and Information Sharing Applications

I was also interested in communication (Figure 3.4) and information sharing (Figure 3.5) habits. As result of the growing prevalence of social networking [Morris 2013], I wanted to determine the degree to which such technologies are utilized for daily work routines. In a multiple choice grid, respondents could select (on a 5-point Likert scale) the frequency of technology usage in times per day. In addition, respondents were able to extend the provided list by adding software applications not listed yet.

The results in figure 6 show the importance of face-to-face communication and established remote communication technologies, i.e., phone and e-mail. This is in line with other studies that identified communication technologies that are part of the everyday work as means to realize CIR [Morris 2013]. It is noticeable that academic social networks seem to play only a small role in enabling communication between colleagues.

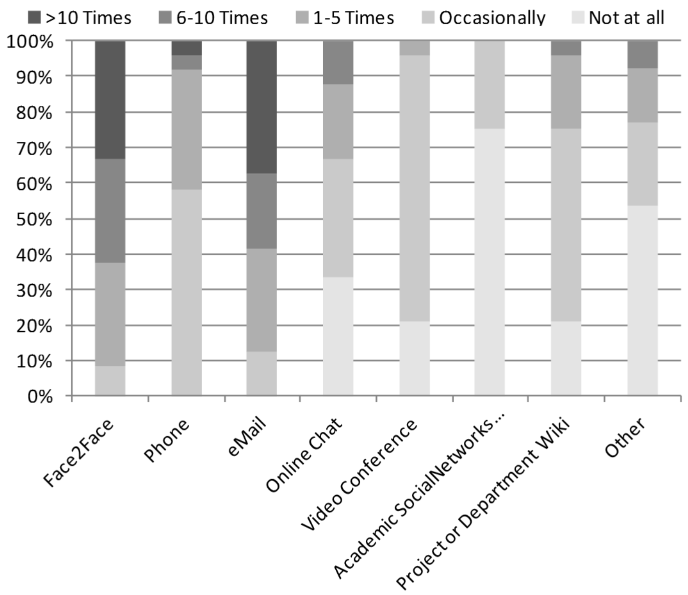



**Figure 3.4:** Frequency of use of various communication technologies

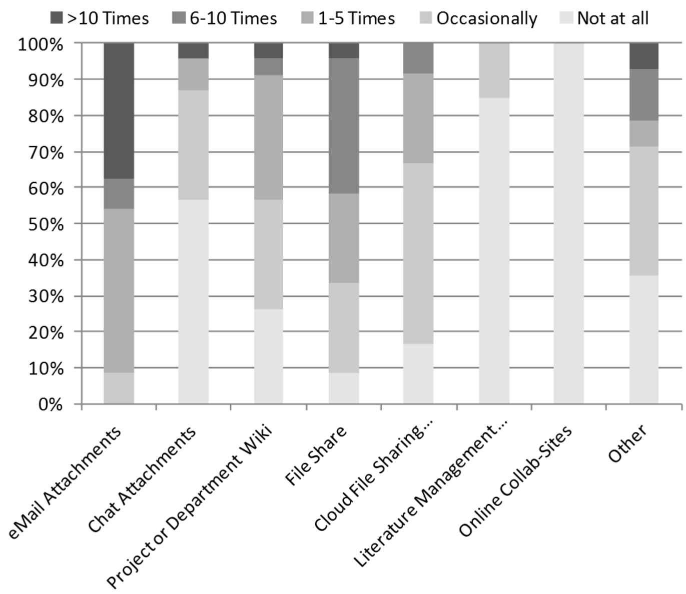



**Figure 3.5:** Frequency of use of different technologies for data and information sharing

Figure 7 depicts technologies for realizing data and information sharing utilized by the respondents. A predominance of e-mail attachments and the usage of file shares can be found. In contrast to this, integrated group support in literature management systems as well as online collaboration sites are rarely in use. A large list of additional software applications (*"Others"*) has been named by respondents which included *Google Drive*, version control systems (namely *GIT*), *Streamworks*, and *SAPmats* (each specified twice). Furthermore, *AeroFS*, *Teambox*, and *Adobe Connect* have been added by respondents. This particularly large number of technologies used for realizing collaborative activities indicates a very heterogeneous collaboration environment where each team member uses his personally preferred software applications.

#### 3.1.2.5 Finding a Partner

I wanted to learn more about how respondents identify colleagues that are expected to be helpful in answering their questions. I asked two questions:

1. *How did you know who might be able to help you*?

2. *How did you contact the person you asked for help?*

I provided an optional free-text field for answers. Twelve respondents provided insights on this process. Some answers showed that colleagues are predominantly approached only if after first clarification using Web-based search wasn't satisfying or helpful:

> *"I try to Google the issue [...]. If that's unsuccessful, I personally contact the colleagues who have experience with that [topic]. I explain my concrete scenario and ask them for help. Sometimes, they don't know the solution but give some new input where to look for."*

After analyzing all answers, I identified three categories of approaching colleagues when looking for help:

1. Random contacting: Respondents ask colleagues without knowing whether they can provide the required information or not (e.g. "*Asking around in the team*", "[*asking*] *whoever is closest*").

2. Specific contacting: Based on a personal social network and an awareness of the qualifications of their team mates, colleagues are directly approached (e.g. *"I asked another biologist who is well versed with [the topic] and has demonstrated that in many fields."*, "[*I asked*] *colleagues who have a longer research experience and/or better background knowledge* [...]").

3. Expert searching: An attempt is made to identify potentially helpful colleagues by looking at the University/research group websites.

Typical ways of contacting colleagues include e-Mail, chat or personal contact with face-to-face communication (i.e., "*[I] went to their office*").

3.1.2.6 Limitations

The demographic targeted by this survey was characterized by high academic degrees and a high experience in research collaboration. Respondents were residents in Germany. Additionally, the relatively small number of respondents might limit the significance of this study. The data I report can probably not be generalized beyond this demographic.

### 3.1.3 Discussion and Design Implications

In line with other research, e.g., [M. Hertzum and Pejtersen 2000], my results confirmed that CIR often also involves looking for informed people. I identified three approaches of identifying a potentially helpful colleague, i.e., expert search, random and specific contacting. The results indicate that collaboration could become more efficient, if team members could better identify colleagues who might be helpful regarding their questions and problems. The task of expert search, however, has recently been addressed in [Engel 2015] and, thus, is not in focus in this dissertation.

The results of my study also indicate that nowadays, collaboration is performed in a heterogeneous environment. That is, it must be assumed that team members use their own personal configuration of software applications for the different collaboration and information activities (i.e., communication, collaboration, data and information sharing, seeking and searching, and result management). This configuration is based on personal preferences, work habits, and the special needs of the team members (e.g., thematically specialized digital libraries). Also in line with previous studies (see section 2.3), collaboration can be observed in all stages of the search process.

As design implications, the results indicate that a coupling of software applications used in everyday work routines represents a necessity for an environment supporting CIR. Instead of providing communication and information sharing means integrated in one system, connecting to external tools and mediating between the co-searchers seems to be a promising way. This might also allow CIR support systems to evaluate the mediated information to infer team support, such as suggestions of potentially beneficial information activities to be performed by team members. In this dissertation, suggesting an information activity to each member of a team is considered a ***Collaboration Strategy***. (Please note that the concept of collaboration strategies is formally defined in the next chapter.)

## 3.2 Conceptual System Model for CIR Support

In this section, the findings of the conducted user study are used to develop and compare the conceptual differences of system models that were designed to support CIR. This is outlined in section 3.2.1. Afterwards, this section discusses the general conditions and requirements that need to be considered when developing CIR support for a team of searchers (see section 3.2.2). The conceptual system model that has been identified as particularly relevant for this dissertation is integrated into a general process model that provides the layout for an environment enhanced by CIR support.

### 3.2.1 Conceptual Differences of CIR System Models

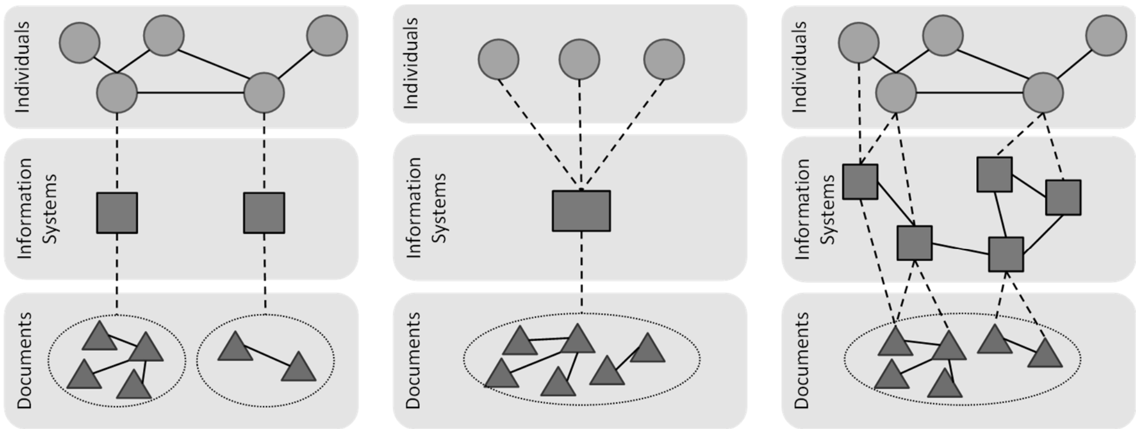
Figure 3.6 depicts the conceptual differences between system models designed to support collaborative work and information activates. This covers the different settings considered in previous research approaches in the area of CIR. Figure 3.6a captures settings identified by empirical studies, such as the ones presented in [Morris 2008] and [Twidale et al. 1997]: It is common for people working together in teams on a project to using IS and IR systems in parallel and to collaborate face to face and through generic communication and data-exchange applications. This is captured in Figure 3.6a where we see a schematic depiction of loosely coupled use of information systems with communication between users performed outside the information system. Moreover, the depiction of Figure 3.6b captures existing prototype CIR Support applications (such as the ones described in section 2.5) that integrate collaboration functions (mediated either in the front-end or back-end) into a central information system, but such systems have not been widely adopted.

In my study, however, I identified that nowadays, knowledge workers utilize a wide range of collaboration support and information systems to meet their daily work demands and satisfy their work-relevant information needs. As a result of my findings, I proposed to model the collaborative work and information environment as a heterogeneous network that consists of three layers [Böhm et al. 2014b]. The elements of this heterogeneous network, depicted in Figure 3.6c, are as follows (from top to bottom):

1) A network of individuals that collaborate during work task performance,

2) A set of utilized information systems and collaboration support tools, and

3) The documents, exchanged, shared, organized, and managed by the utilized information systems and collaboration support tools (see Figure 3.6c).




(a) (b) (c)

**Figure 3.6:** Conceptual differences of CIR system models **[Böhm et al. 2014b]**

### 3.2.2 Process-Model of CIR Support in Heterogeneous Environments

Recent empirical studies showed that, in professional practice, information systems and interfaces designed for individual usage are utilized in collaborative work. The user study presented in section 3.1 supports these findings and states users' habits more precisely. Whereas CIR prototype systems often provided users with functions to support collaboration, for example awareness information about co-workers information activities visualized in the front-end, these systems are not widely adopted. Therefore, such functions are not in focus of this dissertation. In real world settings, however, collaborative users perform their information activities loosely coupled, synchronously or asynchronously and they employ software applications of their daily work routines to realize collaboration [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013].

This has two main consequences for collaborative searchers:

First, when people search for information to satisfy a shared information need, they use traditional search engines and interfaces designed for solitary usage. Hence, user support functions implemented by these information systems concentrates on individual rather than on team level.

Second, the above mentioned studies [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013] on CIR practices showed that team members work individually and that they synchronize their work via loosely coupled communication. Hence, when team members search to satisfy the same information need, they often use the same or very similar query terms [Foley and Smeaton 2009] which is likely to result in highly similar ranked lists returned by the search engine. This may lead to less coverage and less productivity due redundant work [Morris 2007].

To account for these consequences, this section presents a process-model that consists of three phases with the objective to facilitate the CIR effectiveness in a heterogeneous environment. The process-model, depicted in Figure 3.7, covers the following phases:

I) **Tracking Phase.** Collaborative search tasks are performed in a heterogeneous collaboration environment of networked users, information systems, and information objects. This conceptual system model (depicted in depicted in Figure 3.6c) is integrated as a basis into the process-model of CIR support (lower part of Figure 3.7). In the tracking phase, the aim is to track what actually happens during the CIR process. The information activities of team members (performed within this environment) can be logged and stored in an appropriate repository. This logging of users' Information Dialog contexts has previously been proposed as prerequisite for providing user support during the performance of IR tasks [Landwich 2012].

II) **Evaluation and Suggestion Phase.** In this phase, the possible courses of action of team members are explored and evaluated. To this aim, the logged data (depicted in the middle part of Figure 3.7) serves as basis to identify user-specific parameters, also called the user profile [Gauch et al. 2007]. Previous research in the area of IR used such user profiles for tasks such as search personalization [Bennett et al. 2012], predicting future interests [White et al. 2010] or query categorization [Cao et al. 2009], but only considering a single actor. In my scenario, however, I aim at balancing team work by suggesting specific information activities to team members. The profiles of team members together with their current activities (e.g. issued query) are used to come up with optimum suggestions for collaboration strategies (that are formalized in the next section). These suggestions may include, for example, query terms for expansion or document sets for inspection, and aim to consider the trade-off between relevance of a suggestion for a particular user and the resulting progress and effort for the whole team.

III) **Activity Phase.** Finally, the suggestions are provided to the team members who will engage into the corresponding information activities. Figure 3.7 illustrates this by integrating the Information Dialog developed in [Landwich, Klas, et al. 2009] to describe an individual's information searching process (see also section 2.6.1). To reach effective collaboration, it is assumed that team members accept the suggestions. (This phase merges with the Tracking Phase so that the process cycle starts over again.)

The objective of the next section is to develop and formalize the underlying ranking principle for such suggestions in collaborative environments. This is a first attempt towards providing the foundation for CIR Support services in heterogeneous collaborative environments as outlined in Figure 3.7.

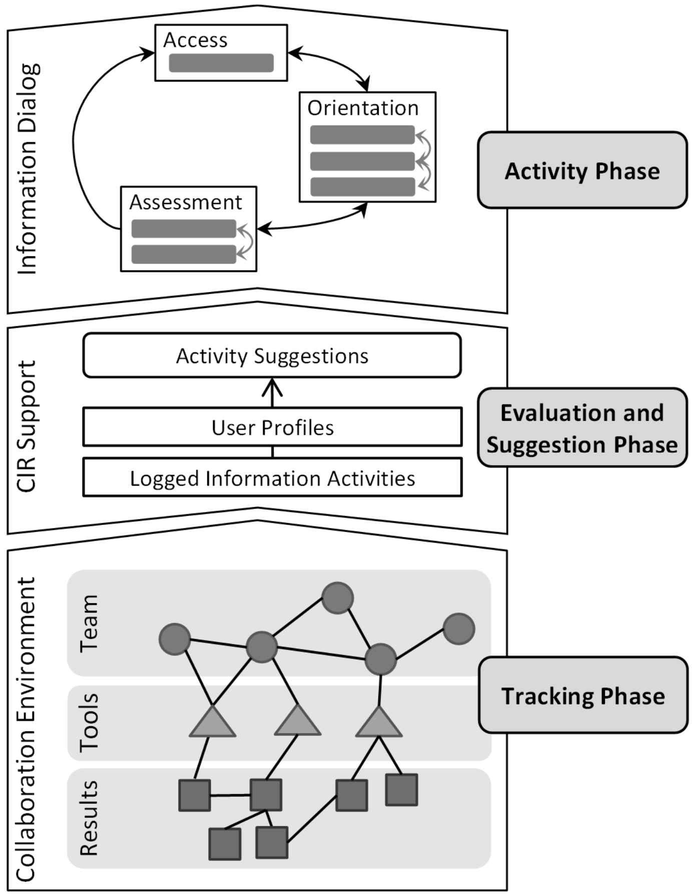



**Figure 3.7:** Process model for supporting a team during CIR

## 3.3 Towards a Ranking Principle for CIR

Ranking is the central problem for many applications of IR and defines the order in which the single items of a search result are presented to the user. In the task of ranking, given a set of items, one utilizes a ranking function to create an ordered list of the items. The relative order of items in this list may represent their degrees of relevance, preference, or importance, depending on application constraints. In this dissertation, these application constraints are given by the characteristics of collaborative search task performance (see section 3.3.3).

In this section, research question *RQ2* (see chapter 1) is approached by introducing a probabilistic model which considers the probability that a document is relevant and non-redundant. Documents are ranked on the basis of two probabilities: the probability of relevance of a document with respect to the searchers' information need and the probability of another team member also discovering this particular document. The approach taken is justified by decision theory, i.e., it aims at minimizing expected costs for the team (see section 3.3.5).

### 3.3.1 Objective of the Ranking Principle

Early IR approaches were based on exact-match models, such as the Boolean model, that identified documents based on an information need expressed using, e.g., Boolean logic [Salton and McGill 1983]. In those models, documents are retrieved if they completely fulfill the issued query. Thus, the query response is a set of documents without any order. Therefore, each document within the result set must be assumed to be as important as any other retrieved document. Several extensions of the Boolean retrieval model as well as alternative models have been proposed [Salton and McGill 1983], each of them aiming to assign scores to retrieved documents. However, in this chapter I focus on probabilistic models because optimal retrieval has been precisely defined only for probabilistic models where the optimality of the ranking strategy can be proofed formally [Fuhr 1992].

I consider documents ranked in sequential order. According to Gordon and Lenk [Gordon and Lenk 1991], user satisfaction is maximized if documents are returned in an order which minimizes the number of documents to be inspected by users to satisfy their information need. Such a sequential ranking of documents is related to sorting a list of documents. Ranking is performed according to a ranking criterion which expresses the order of two given documents (i.e., I assume comparable documents). Ranking of a whole list of documents can then be realized using, e.g., the bubble-sort algorithm whose correctness can be proofed by induction (see [Knuth 1998]).

The objective of this section is to develop such a ranking criterion which expresses the cost-optimal order of two given documents in a collaborative search session.

### 3.3.2 Foundations of Probabilistic Ranking

IR is characterized by an inherent uncertainty [Fuhr 1992] because an IR system can only have a limited understanding of features like documents, queries as well as the relevance relationship between them. This is because it operates on representations of the information needs and original documents. A query formulation cannot be assumed to precisely represent an information need. There is also no clear procedure that decides whether or not a document is an answer [Fuhr 1992]. The complexity of an information retrieval task also results from this uncertainty. Generally speaking, essential features are impossible to quantify precisely; this is especially true when human behavior is involved (e.g., transforming the information need into a query). Instead, probabilistic models aim at abstracting those essential features and aggregate them in some systematic way leading to model components with manageable numbers of parameters.

Probabilistic models in IR are often characterized by introducing an explicit variable $\mathfrak{R}$ that abstracts the relevance (or non-relevance) of a document. This variable is not observable directly and hence, its value is uncertain, i.e., $\mathfrak{R}$ is a random variable. A probability distribution represents the document's estimated probability of relevance with respect to a query by encapsulating them in the conditional probability $P(\mathfrak{R}|d,q)$ [Robertson and Zaragoza 2009]. However, the process of ranking considers documents w.r.t. a fixed query. It is common to express the probability of document relevance by $P(\mathfrak{R}|d)$. The order in which the documents are presented to the user is defined by the PRP:

> *"If retrieved documents are ordered by decreasing probability of relevance on the data available, then the system's effectiveness is the best that can be obtained for the data."* [Robertson and Zaragoza 2009]

However, the process of estimating the probability of relevance may not consider only the issued query $q$, but also additional observable data, such as a user profile and the search task's context, depending on the particular IR application [Lavrenko 2010].

Decision theory has often been used in IR research for coming up with solutions or criteria for various IR tasks. This covers, e.g., database selection in networked IR [Fuhr 1999], the justification of the PRP [Robertson 1977], and the development of the Probability Ranking Principle for Interactive IR (iPRP) [Fuhr 2008].

Decision theory is concerned with determining which decision, from a set of possible alternatives, is optimal. The decision is characterized by several alternatives and the consequences resulting from a choice are imperfectly known, i.e., the decisions are made in the face of uncertainty [North 1968]. Each decision will incur costs that are quantified by a loss function $\mathcal{L}$ which depends on the true state of nature $\omega_k$. States of nature are events that are likely to occur and over which the decision maker has no control.

In the context of IR, states of nature may be the relevance or non-relevance of a document, i.e. $\omega_k \in \{'relevant', 'not\ relevant'\}$. The resulting expected costs are usually defined as the sum of costs multiplied by the probability of that cost occurring:

$$EC(d) = \sum_k \mathcal{L}(\omega_k)\, P(\omega_k|d) \tag{3.1}$$

### 3.3.3 Assumed Characteristics of Collaborative Search Tasks

Robertson defined the following assumption for the PRP to make it hold unequivocally [Robertson 1977]: '*The relevance of a document to a request is independent of the other documents in the collections*'. Early analysis showed that in real world search scenarios, this assumption may not hold [Gordon and Lenk 1992]. Modifications of the PRP have been proposed recently that consider a dependence between documents, e.g. [Fuhr 2008]. However, as the PRP is the most prominent ranking principle, I used it as template.

I define the following additional assumptions for a collaborative search task.

- Team members work loosely coupled [Patel and Kalter 1993]: Participants use independent applications which do not integrate collaboration services, such as awareness information of team member's activities [Morris 2013; Kelly and Payne 2014; Böhm et al. 2013]. They communicate infrequently using one-shot information exchange to copy the latest progress-state among another, i.e., team members act independently.
- Team members have a common understanding of the shared information need and the (electronic) information sources to use. Users with profound domain knowledge generally use a more specific vocabulary [White et al. 2009] and thus, they construct own queries. However, these queries address the same, shared information need.
- Focus is on recall and productivity of the team: As a typical characteristic of professional search, the aim is on identifying as much relevant information as possible (i.e., maximum recall) [Joho et al. 2010]. At the same time, this should

be done at minimum costs (e.g. financial costs/time required), i.e., minimum redundancy.

### 3.3.4 Set-Theoretic Representation of Collaborative Information Activities

My approach is based on modeling a collaborative search task by describing the document sets, such as (electronic) sources, retrieved, inspected, and assessed documents, associated with team members over the course of search. In this way, the model refers to the system's representation of information and information needs, as it is common in probabilistic models [Fuhr 2008] [Fuhr 1992].

I apply the general collaboration framework by Baeza-Yates and Pino who described a collaborative task to be performed by a team $T$ consisting of $N$ team members and the task to be divided in $L$ stages [Baeza-Yates and Pino 1997]. Team members $\tau_i \in T$ perform iterative search sessions individually and the relevant search results of all team members are aggregated. I use the Information Dialog developed by Landwich, Klas et al. to describe an individual's information searching process [Landwich, Klas, et al. 2009]. Different document sets and activities are associated with these stages[2]:

1) Activities of Access produce a result set of documents from a given source as response to query. The elements of this set have been determined by the IR system based on an estimated probability of relevance, i.e., system relevance [Saracevic 2007].

   **Definition 3.1:** ***Query Result.*** Let $D$ be the set of documents contained in an information source. The result in response to a query $q_i$ issued by team member $\tau_i \in T$ is defined as sub-set $R(q_i) \subseteq D$. Let $\chi$ be real cut-off value. Documents belong to a query response $R(q_i)$ if their probabilities of relevance are at least that of irrelevance and exceeding the cut-off value, i.e. $R(q_i) = \{d | P(\Re|d) > P(\bar{\Re}|d) \wedge P(\Re|d) > \chi\}$.

2) Activities of Orientation create result sub-sets which reach the field of vision of the user. For example, a user might decide to scroll through the result list until a

[2] Please refer to [Landwich 2012] for detailed definitions of the document sets associated with the stages of the Information Dialog. For the sake of simplicity, I only introduce the minimum amount of formal definitions required for the approach introduced in this chapter.

certain rank or request another page of the result set. In both cases, the user captures the document visualization (e.g. Rich Snippet[3] format) cognitively.

**Definition 3.2:** ***Viewed Result.*** Let $R(q_i)$ be a query response. The set of documents viewed by team member $\tau_i \in T$ is defined as sub-set $V(q_i) \subseteq R(q_i)$.

Finally, a sub-set of documents is selected by the user for inspection. That is, the user might request the abstract or full-text of a document to eventually read it.

**Definition 3.3:** ***Inspected Result.*** Let $V(q_i)$ be sub-set of document of a query response a user has requested and viewed. The set of documents inspected by team member $\tau_i \in T$ is defined as sub-set $I(q_i) \subseteq V(q_i)$.

3) Activities of Assessment identify useful documents for the task at hand. For example, the inspection of a document allows the user to verify (assess) the affective relevance [Saracevic 2007] of it.

**Definition 3.4:** ***Accepted Result.*** Let $I(q_i)$ be set of inspected documents of a query response. The set of documents verified as relevant by team member $\tau_i \in T$ is defined as sub-set $A(q_i) \subseteq I(q_i) \subseteq V(q_i) \subseteq R(q_i) \subseteq D$.

The main characteristic of a collaborative search task is that it involves multiple users aiming at collaboratively solving a shared information need [G. Golovchinsky et al. 2008]. Each team member individually performs several cycles of the Information Dialog described above. For constructing a schematic visualization of such search sessions, Venn diagrams of documents-sets have been used as means in [Landwich, Klas, et al. 2009] as well as in [Hansen 2011]. In the latter work, those visualizations were also used as means to visualize collaborative information activities. Figure 3.8a depicts an example of such a general schematic visualization. We see information activities of two collaborative users in timely order. Each of them issued a query, viewed, inspected, and assessed results.

[3] Rich Snippet: https://developers.google.com/structured-data/rich-snippets/ retrieved 25.01.2016

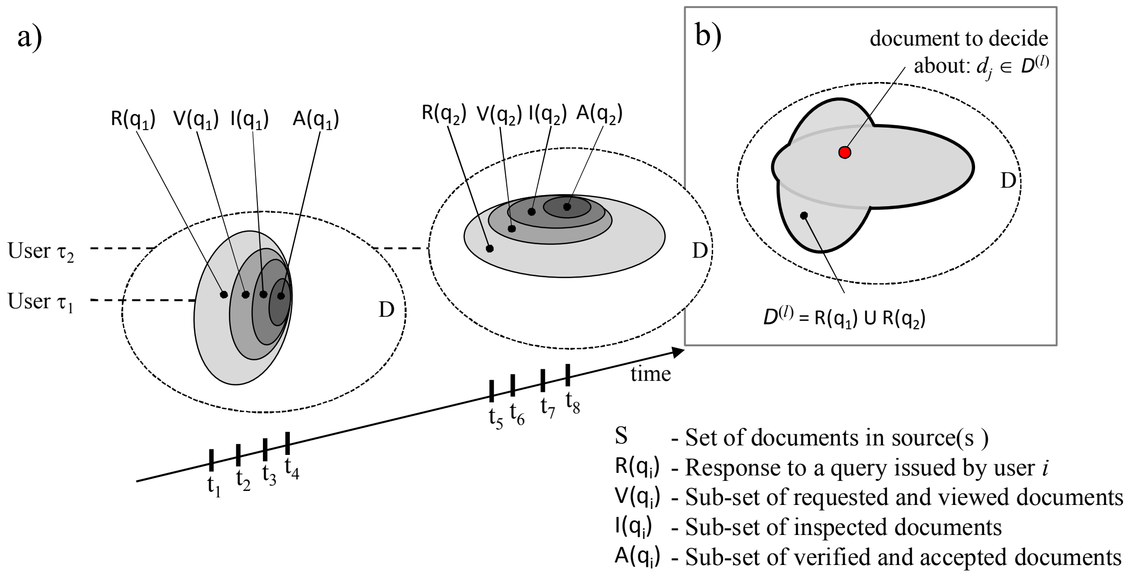



**Figure 3.8:** Set-theoretic representation of collaborative information activities of two users

When team members search to satisfy the same information need, they often use the same or very similar query terms [Foley and Smeaton 2009] which is likely to result in highly similar ranked document lists returned by the search engine. Figure 3.8b depicts the projection of the query responses associated with both team members. As can be seen in Figure 3.8b, the document sets retrieved by team members can overlap. This may lead to less coverage and less productivity due redundant work [Morris 2007].

Under this premise, a nuanced approach is required which aims at increasing the chances that team members will identify relevant documents (that otherwise would be lost) by increasing the teams' coverage of the information space, i.e., the document corpus *D*. In the example of Figure 3.8, the union of documents $D^{(l)} = R(q_1) \cup R(q_2)$ at time point *l* represents the basis for my ranking approach.

From a system's point of view, the question arises, how the system should rank document $d_j \in D^{(l)}$ for each team member. Based on the PRP, the rank will be defined based on the estimated relevance of a document. However, in a collaborative scenario, my hypothesis is, that not only estimated relevance, but also potential information activities of the other team members and resulting redundancy need to be considered to obtain an optimal, i.e. cost minimizing, ranking. However, in the setting considered in this dissertation, information activities of team mates are not observable.

In order to formalize this adequately, I introduce the following definitions.

**Definition 3.5:** ***Shared Result Set***. Let $T$ be a set of $N$ team members. A shared result set at time-point $l$ is defined as union $D^{(l)} = \bigcup_i^N R(q_i)$.

**Definition 3.6:** ***Coverage***. Let $Q_T$ be the set of queries issued by all team members within a collaborative search session. The union of documents inspected by all team members is defined as $Coverage(T) = \bigcup_{q \in Q} I(q)$.

**Definition 3.6:** ***Effort***. Let $Q_T$ be the set of queries issued by all team members within a collaborative search session. The sum of documents inspected by all team members is defined as $Effort(T) = \sum_{q \in Q} |I(q)|$.

**Definition 3.7:** ***Relevant Coverage***. Let $Q_T$ be the set of queries issued by all team members within a collaborative search session. The union of relevant documents found by all team members is defined as $RelCov(T) = \bigcup_{q \in Q} A(q)$.

### 3.3.5 A Cost Model for Collaborative Search

In decision theoretic models, abstract costs are introduced that may cover computing or response times, or the effort required by a user to do his or her job [North 1968] [Fuhr 1999]. For the justification of the classical PRP, Robertson [Robertson 1977] introduced a simple loss function that defined costs associated with the decision as to whether or not to retrieve a document depending on its expected relevance with regard to a searcher's information need. Retrieving a relevant document may result in the abstract costs denoted with $B$, retrieving a non-relevant document resulted in costs denoted with $C$. The only assumption made was $B < C$ which allowed to proof the cost-optimality of the PRP (see also [Fuhr 1992] and [Bookstein 1983]).

In my model, decisions are made about suggesting documents to a team member who is assumed to inspect and assess them. Each decision may result in different costs depending on the (non-observable) true state of nature. For my attempt towards an initial cost model for CIR, however, I wanted to maintain the clarity of Robertson's approach and adjust the corresponding cost model only slightly to make it applicable to CIR.

In a collaborative scenario, information activities of team members also influence the true state of nature of a document because the same document $d_j \in D^{(l)}$ could be inspected multiple times. Thus, a relevant document does not necessarily translate into a benefit $B$ (negative quantity), but in wasted efforts $\bar{B}$ (positive quantity) due to redundant work if a team member discovered it already (i.e., relevant-but-redundant documents).

My probabilistic model for CIR is characterized by introducing a second probabilistic parameter $\delta_{\neg i,j}$ which is an estimate about information activities of other team members. Besides the relevance relation between document and information need (expressed by the formalized query $q_i$ issued by team member $\tau_i$), documents may relate to other team mates $\tau_{\neg i}$ according to $\delta_{\neg i,j}$, that is, how likely it is that another team mate $\tau_{\neg i}$ discovered the corresponding document during the course of search. Thus, in my model, state of nature (in addition to relevance) also depends on the likelihood that the document is redundant. The following list summarizes resulting costs that depend on the true state of nature:

$$\mathcal{L}('relevant' \wedge 'not\ discovered') = B$$

$$\mathcal{L}('relevant' \wedge 'discovered') = \bar{B}$$

$$\mathcal{L}('not\ relevant') = C$$

Let $\rho_{i,j}$ be the probabilistic parameter describing the probability of document $d_j$ being relevant. Furthermore, let $\delta_{\neg i,j}$ be the probability that document $d_j$ has been discovered by another team member $\tau_{\neg i}$, i.e., document $d_j$ is redundant. (Please see the next section for a detailed definition of the underlying event space.) Team members are assumed to act independent so that resulting events of relevance and redundancy are expressed by the products of the probabilistic parameters:

$$P('relevant' \wedge 'not\ discovered'|d_j) = \rho_{i,j} \cdot (1 - \delta_{\neg i,j})$$

$$P('relevant' \wedge 'discovered'|d_j) = \rho_{i,j} \cdot \delta_{\neg i,j}$$

$$P('not\ relevant'|d_j) = (1 - \rho_{i,j})$$

We can summarize the expected costs incurred by suggesting a document by building the sum of the defined losses and associated probabilities of occurrences according to equation 4.1 (of section 3.3.2).

$$EC(d_j) = \rho_{i,j}(1 - \delta_{\neg i,j})B + \rho_{i,j}\delta_{\neg i,j}\bar{B} + (1 - \rho_{i,j})C \qquad (3.2)$$

Equation 3.2 represents a rather simple cost-model, but it does capture the main elements of interest and is similar to the cost-model originally proposed for the justification of the PRP [Robertson 1977]. Equation 3.2 suggests that, in order to minimize the expected costs, a system should allocate documents that satisfy the information need and at the same time have not been discovered by another team member yet.

### 3.3.6 Definition of the Underlying Event Space

The event space underlying most probabilistic models is the Cartesian cross-product of existent documents $D$ and (the universe of) queries $Q$, i.e. $\Omega = Q \times D$. A single element of this event space is a query-document pair $(q_i, d_j)$ [Fuhr 1992]. The probability that a particular document $d_j$ is of relevance with regard to query $q_i$ issued by team member $\tau_{\mathrm{i}}$ is expressed by the conditional probability $P(\Re | q_i, d_j)$. This corresponds to the probabilistic parameter $\rho_{i,j}$ used in equation 3.2 where both, $q_i$ and $d_j$ are observables.

However, in a collaborative scenario, one also needs to consider the set $T \backslash \{\tau_i\}$ of team members who may have issued queries that would eventually lead to the discovery of the same document $d_j$. One can model this by defining $Q_{\neg i} \subseteq Q$ as the set of queries that have been issued by the sub-team $T \backslash \{\tau_i\}$. Each query $q \in Q_{\neg i}$ will create a result set $R(q)$, and from such retrieved results, each team member only views and inspects a sub-set $I(q) \subseteq V(q) \subseteq R(q)$. I define and $I_{\neg i} = \bigcup_{q \in Q_{\neg i}} I(q)$, so that we can consider the event space $\Phi = Q_{\neg i} \times I_{\neg i} \subseteq \Omega$. Because the set of queries issued by the sub-team $T \backslash \{\tau_i\}$ is not observable, I introduce a random variable $\sigma_{\neg i}$ ranging over $Q_{\neg i}$. Moreover, let $\mathfrak{D}$ be a random variable ranging over $\Phi$. The probabilistic parameter $\delta_{\neg i,j}$ can now be defined as $P(\mathfrak{D} \in \Phi | \mathfrak{D} = (\sigma_{\neg i}, d_j) \wedge d_j \in D^{(l)})$.

I present a simple estimation of $\delta_{\neg i,j}$ in section 3.3.9.

### 3.3.7 The cPRP

The cost model introduced in the previous section allows for deriving a ranking principle for collaborative search by making certain assumptions about the relation between the various cost constants. That is, I assume that a non-relevant document will incur approximately the same costs as a redundant document, i.e. $C \approx \bar{B}$. However, a relevant document will always result is lower costs, i.e., $B < C$. This yields the following ranking principle which I denote with cPRP.

**Theorem 3.1**: In order to maximize the productivity of a collaborative search task, an IR system should rank documents according to decreasing values of:

$$\rho_{i,j}\left(1-\delta_{\neg i,j}\right) \tag{3.3}$$

### 3.3.8 Proof of Optimality

The cost-optimality of the ranking principle expressed in equation 3.3 can formally be proofed which is presented below. This proof has the same structure as the proof of the PRP's cost-optimality presented in [Fuhr 1992].

**Proof**: The cPRP is optimal with respect to cost minimization.

Ranking document $j$ before another document $k$ is cost optimal iff:

$$EC\left(d_j\right) \leq EC(d_k) \tag{3.4}$$

This is equivalent to

$$\begin{aligned}\rho_{i,j}(1-\delta_{\neg i,j})B + \rho_{i,j}\delta_{\neg i,j}\bar{B} + \left(1-\rho_{i,j}\right)C \\ \leq \rho_{i,k}(1-\delta_{\neg i,k})B + \rho_{i,k}\delta_{\neg i,k}\bar{B} + \left(1-\rho_{i,k}\right)C\end{aligned} \tag{3.5}$$

This can be simplified to:

$$\rho_{i,j}(B-C) - \rho_{i,j}\delta_{\neg i,j}(B-\bar{B}) \leq \rho_{i,k}(B-C) - \rho_{i,k}\delta_{\neg i,k}(B-\bar{B}) \tag{3.6}$$

Based on the assumption $C \approx \bar{B}$, it applies that $(B-\bar{B})/(B-C) \approx 1$. Because we also assume $B < C$, it applies that $B - C$ is a negative quantity. Thus, dividing the inequality by $B - C$ reverses the inequality and yields:

$$\rho_{i,j} - \rho_{i,j}\delta_{\neg i,j} \geq \rho_{i,k} - \rho_{i,k}\delta_{\neg i,k} \tag{3.7}$$

This is equivalent to:

$$\rho_{i,j}(1-\delta_{\neg i,j}) \geq \rho_{i,k}(1-\delta_{\neg i,k}) \tag{3.8}$$

Inequality 3.8 represents a criterion for bringing two adjacent documents into the cost minimizing order. A whole list of documents can be ordered by applying the criterion iteratively according to, e.g., the bubble sort algorithm whose correctness can also be proofed [Knuth 1998]. This will bring the whole list into an order where the expected costs are minimized.

□

### 3.3.9 An Illustrative Example

As an illustrative example, I briefly demonstrate the application of the cPRP and compare it with the PRP. I do this by discussing a simplified, abstract use-case, which is adopted from that presented by Robertson, to illustrate the application of the PRP in [Robertson 1977]. In my example, two team members aim at satisfying the same, shared information need. They act asynchronously and loosely coupled. There are three documents that will be of interest for this team ($d_1$-$d_3$), and two documents that are not ($d_4$-$d_5$). We suppose that the IR system receives two formalized queries at different, subsequent time points, and both queries are representations of the shared information need. The obvious ranking (according to the PRP) in which the system could present the documents is:

$$L_{PRP} = < d_1, d_2, d_3, d_4, d_5 >$$

This is an optimal ranking in response to the first search request[4]. However, according to the PRP, the response to the second search request would have the same order because the second query addresses the same information need. However, the probability that the first team mate has discovered a document may depend on the document's rank in the result list because real users typically proceed down a rank list only until reaching a specific maximum rank [Jansen et al. 2000], so that the probability of document discovery decreases with increasing ranks. Let $rank(d_j)$ be a function that returns the rank of a document in $L_{PRP}$. So, for the purpose of this example, I simply assume $\delta_{\neg i,j} \propto 1/rank(d_j)$ which describes decreasing probabilities of discovery as the document rank increases[5]. This results in the following ranked list as response to the second query (according to the cPRP):

$$L_{cPRP} = < d_3, d_2, d_1, d_4, d_5 >$$

As one can see, using cPRP-based ranking, the second searcher obtains a list of results that avoids redundant work until a certain number of documents has been assessed by each team member. For example, if both users only examine the first two documents of the lists provided to them, they have reached full satisfaction with regard to their

[4] In response to the first search request, the cPRP-based ranking is equivalent to the PRP-based ranking because the requests are received subsequently and no document could have been discovered before.
5 Please see Table 3.1 for the resulting probabilities

information need (i.e., they found in sum all three relevant documents) and at the same time have only created a little amount of redundant work (i.e., both have inspected and assessed document $d_2$).

This simplified example demonstrated the potential effectiveness of the ranking principle. However, as could be seen, it does not necessarily lead to minimum costs for the team because both team members inspected and assessed document $d_2$. In fact, each team members could have proceeded down the whole ranked list which would lead to maximum redundancy. The latter phenomenon motivated the development of Activity Suggestions (see next section) that suggest a sub-set of documents to be inspected and assessed by each team member to achieve optimum collaboration results.

**Table 3.1:** Summary of probabilities related to a collaborative search example task

| | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ |
|---|---|---|---|---|---|
| Probabilities according to PRP | 0.334 | 0.333 | 0.333 | 0.000 | 0.000 |
| $\delta_{\neg i,j} \propto 1/rank(d_j)$ | 1.000 | 0.500 | 0.333 | 0.250 | 0.200 |
| Probabilities according to cPRP | 0.000 | 0.167 | 0.222 | 0.000 | 0.000 |

The table above (Table 3.1) summarizes the probabilities that led to the second result list. The first row lists the estimated relevance of documents (which corresponds to the level of expected user satisfaction). The second row lists the probability that the first user has discovered the document from a previous result list. Finally, the values of the third row represent probabilities estimations according to the cPRP.

## 3.4 Activity Suggestions for CIR Support

Based on the formal cost model for collaborative search sessions develop in section 3.3.5, a ranking criterion has been derived that can be used for optimum search result ordering (see section 3.3.7). However, in this section I use the underlying cost model to derive a general criterion that describes optimum collaboration strategies in IR, i.e., estimates about which document should be inspected by whom. A CIR support system can suggest those strategies to team members with the aim to facilitate the collaborative performance of IR tasks.

This dissertation introduces the notion of Activity Suggestions that assign a specific sub-set of documents $D_i^{(l)} \subseteq D^{(l)}$ to each team member $\tau_i \in T$ for inspection and assessment. An IR system has a set of options to choose from which can be described by $T \times \mathcal{P}(D^{(l)})$. That is, the IR system may assign some sub-set of $D^{(l)}$, i.e., an element of $\mathcal{P}(D^{(l)})$, to any team member $\tau_i \in T$.

Optimizing the individual's contributions calls for allocating the available documents with respect to the information activities of the team and suggesting an appropriate set of documents $D_i^{(l)} \in \mathcal{P}(D^{(l)})$ to team member $\tau_i \in T$. This is described by a suggestion mapping $s\colon T \times D^{(l)} \rightarrow \{0,1\}$. If a document $d_j \in D^{(l)}$ is suggested to a team member $\tau_i \in T$ the tuple $(\tau_i, d_j)$ is mapped to 1; in case of no suggestion it is mapped to 0. Moreover, I define $M = |D^{(l)}|$. This allows for defining each sub-set $D_i^{(l)}$ as follows:

$$D_i^{(l)} = \{d_j \in D^{(l)} | s_{i,j} = 1\} \tag{3.9}$$

**Definition 3.8.** ***Collaboration Strategy***: A collaboration strategy is any mapping $a\colon T \times D^{(l)} \rightarrow \{0,1\}$ that specifies a sub-set $D_i^{(l)} \subseteq D^{(l)}$ for each team member $\tau_i \in T$, so that $D_i^{(l)} = \{d_j \in D^{(l)} | s_{i,j} = 1\}$.

If one assumes that the sub-sets $D_i^{(l)}$ are known, the expected costs can be formulated considering the whole team $T$ using equation 3.2 as follows

$$EC(D_1^{(l)}, \ldots, D_N^{(l)}) = \sum_{i}^{N} \sum_{d_j \in D_i^{(l)}} EC(d_j) \tag{3.10}$$

However, for the purpose of involving the suggestion mapping into this equation, i.e., formulate the expected costs $EC(D_1^{(l)}, \ldots, D_N^{(l)} | s)$ that are conditioned by the suggestion mapping, I furthermore introduce the following two constraints.

I consider that real users request only a limited number of result-pages and inspect only the top-$K_{max}$ result records [Jansen et al. 2000]. My model reflects this by introducing a user-specific parameter which I denote with $K_i$. This parameter describes the maximum rank that a user is willing to proceed down a ranked result list. Furthermore, introducing an additional constraint allows assuring that $\delta_{\neg i,j}$ is zero: This is achieved by limiting the number of times a document is suggested to users to one (to avoid redundant work). This allows describing the overall costs for a team involving the suggestion mapping $s$:

$$EC(\,D_1^{(l)}, \ldots,\, D_N^{(l)}|s) = \sum_i^N \sum_j^M s_{i,j}\left[\rho_{i,j}B + (1-\rho_{i,j})C\right] \quad (3.11)$$
$$\text{subject to } \textstyle\sum_j^M s_{i,j} \leq K_i, \forall i \text{ and } \sum_i^N s_{i,j} \leq 1, \forall j$$

### 3.4.1 Optimum Suggestions

I now want to develop a criterion (or rule) that ensures that the costs resulting from the suggestion mapping $s$, as given by equation 3.11, are minimized. To this aim, I take into account the assumption that costs incurred by a relevant document are less than those incurred by non-relevant one, i.e. $B < C$. This leads to the conclusion that the expected costs $\rho_{i,j}B + (1-\rho_{i,j})C$ are strictly monotonically decreasing with $\rho_{i,j} = [0,1]$. Thus, minimizing the expected costs corresponds to the following objective function and constraints that represent an integer linear program (ILP):

$$\min EC(\,D_1^{(l)}, \ldots,\, D_N^{(l)}|s) \Leftrightarrow \max \sum_i^N \sum_j^M s_{i,j}\rho_{i,j} \quad (3.12)$$
$$\text{subject to } \textstyle\sum_j^M s_{i,j} \leq K_i, \forall i \text{ and } \sum_i^N s_{i,j} \leq 1, \forall j$$

The following theorem represents the optimum criterion for collaborative search.

**Theorem 3.2:** In order to maximize the productivity of a collaborative search task, an IR system should allocate documents to team members according to:

$$\max \sum_i^N \sum_j^M s_{i,j}\rho_{i,j} \quad (3.13)$$
$$\text{subject to } \textstyle\sum_j^M s_{i,j} \leq K_i, \forall i \text{ and } \sum_i^N s_{i,j} \leq 1, \forall j$$

### 3.4.2 Model Analysis

In this criterion (equation 3.13), the suggestion mapping $s_{i,j}$ represents the unknowns of the ILP to be determined and $\rho_{i,j}$ represents the user-specific relevance probabilities that need to be estimated beforehand.

Considering the cut-off value $\chi$ introduced in section 3.3.2, defining $\rho_{i,j} = P(\Re|q_i, d_j) - \chi$ assures that the solution of the ILP considers only documents with a probability of relevance greater than the cut-off value $\chi$. If we assume that the cost constants of my model have been quantified, parameter $\chi$ can be estimated for any arbitrary document $d_j$, for which $\exists i: s_{i,j} = 1$ is true (i.e., it is suggested to some team member) as follows: One requires that the maximum expected costs incurred by a document is zero. (Recall that the minimum costs are $B$, since I defined $B < 0$.) One can write this assertion as $\rho_{i,j}B + (1 - \rho_{i,j})C = 0$ with $\rho_{i,j} = -\chi$. This yields $\chi = \frac{C}{C-B}$. So, we see that $\chi$ is uniform among documents. Also, since I already introduced the assumption $C > B$, one can conclude that $C - B$ is a positive quantity and thus $\chi = ]0,1]$.

For estimating $P(\Re|q_i, d_j)$, IR research provides a large body of knowledge covering many approaches for computing query expansion terms and re-ranking of results, each with the aim of increasing the quality of search results towards resolving the actual information need. As an example, research in the field of search personalization, see for example [Bennett et al. 2012] or [Bouhini et al. 2014], provides useful approaches that are based on building a user-profile from the search history and incorporating this profile into the retrieval function.

However, for my model the user-specific parameter $K_i$ still needs to be estimated but this could also be done based on the user's history, e.g., by using the average number of documents the user has assessed per query in the past.

With regard to the illustrative example presented in section 3.3.9, one possible solution of the corresponding ILP and a resulting suggestion mapping is represented as matrix as follows:

$$s_{i,j} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

In this matrix, rows represent team members and are indexed with $i$, and columns represent documents and are indexed by $j$: This is an optimal collaboration strategy because (in contrast to the ranking result presented in section 3.3.9) the suggestion mapping does not result in redundant work since each document is mapped only once (in particular, document $d_2$ is assigned only once).

## 3.5 Discussion of Activity Suggestions and their Limitations

In this section, I mainly discuss the formal criterion for optimum collaboration strategies denoted with Activity Suggestions. I do this because the next chapters will demonstrate the practicability of the developed formal criterion by a prototypical implementation (see chapter 4) and exemplify the application of Activity Suggestions by means of search result division among team members in a quantitative evaluation (see chapter 5).

Based on the data gathered using an empirical user study, an informal system model has been developed that outlined the technical environment in which CIR is typically performed. To be able to formalize CIR support in such environments, a process model of CIR support has been developed afterwards. This chapter continued with the formalization the ranking principle underlying this CIR support. To this aim, I developed a cost-model that is based on describing the information activities of single team-members using the Information Dialog. Due to the assumption that team members perform their information activities independently, I introduced two probabilistic parameters describing document relevance as well as document redundancy. Moreover, these parameters have been integrated into a cost model for describing collaborative search sessions.

From the cost model introduced in this chapter, I derived the notion of Activity Suggestions that represent a formal criterion that can be used for determining optimum collaboration strategies of teams, e.g., result division among team members. It is important to note that the developed criterion is ***declarative***, i.e., it describes how an optimum result division is characterized, but it does not explicitly define how this optimum is reached or computed. This is in contrast to previous ***imperative*** approaches, where a scoring function [Jeremy Pickens et al. 2008] or algorithm [Soulier et al. 2013] were hypnotized to result in better IR performance for a team.

It is interesting to note that, although the cost function introduced in section 3.3.5 depends on several variables, the derived criterion has a rather simple structure and can be computed easily using a numeric solver for ILPs.

A second interesting issue is that, if one assumes a team size $N = 1$ and a mapping $a\colon D^{(l)} \rightarrow \{0,1\}$, the PRP can be derived as a special case of my optimum criterion:

$$\max \sum_{j}^{M} a_j p_j \qquad (3.14)$$

subject to $\sum_{j}^{M} a_j \leq K_1$

This equation represents an alternative formulation of the PRP: For each rank, or for each given number of documents $K_i$ requested by a user, respectively, documents with a maximum probability of relevance are allocated. Hence, my optimum criterion represents a generalization of the well-known PRP, but also includes the same limitations, such as assumed independence between documents.

However, with Activity Suggestions formalized in this chapter, the question arises how the developed theory can be integrated into existing IR technology. To this aim, the next chapter presents a prototypical implementation of Activity Suggestions which demonstrates the practicability of the developed formal criterion.

# 4 Implementing CIR Support

This chapter presents the prototypical implementation of Activity Suggestions which eventually shall be employed in a quantitative evaluation realized as computer simulation of collaborative search tasks in two professional domains.

Recently, simulations in IR research have attracted a lot of attention and became topic of interest [Azzopardi et al. 2011]. Simulations enable researchers to conduct carefully designed and controlled experiments to obtain answers to research questions regarding the IR process. Typically, simulations are designed to replicate and mimic different aspects of the IR process and to perform "*What-If*" experiments, that is, experiments that allow for exploring a hypothesis about different courses of interaction or different IR models [Azzopardi et al. 2011].

Figure 4.1 depicts the main high-level elements within a simulation schematically (adopted from [Azzopardi et al. 2011], page 37). These include a model of the user (i.e., the simulated user) and a model of the interface. The implementation of the elements that are described in this section are encapsulated using a dashed line in Figure 4.1.

Whereas the user model encapsulates real users and their behavior, such as performed information activities and responses to events and how they assess document relevance, the interface model provides an abstraction of the IR system's interface itself and exposes the main functionality of the system to the simulated user. A simulated user represents an instance of the user model.

In Figure 4.1, the entity which's effects on the (collaborative) IR process shall be evaluated using simulation is depicted as a component ("*Agent to be evaluated*") within the IR system.

This chapter provides a description of the simulation software for CIR tasks that has been developed for and used in this dissertation for experimentation and evaluation purposes. This chapter covers the definition of the user model as well as the interface model. Additionally, the implementation strategy for implementing Activity Suggestions as an additional agent within a general IR system is outlined.

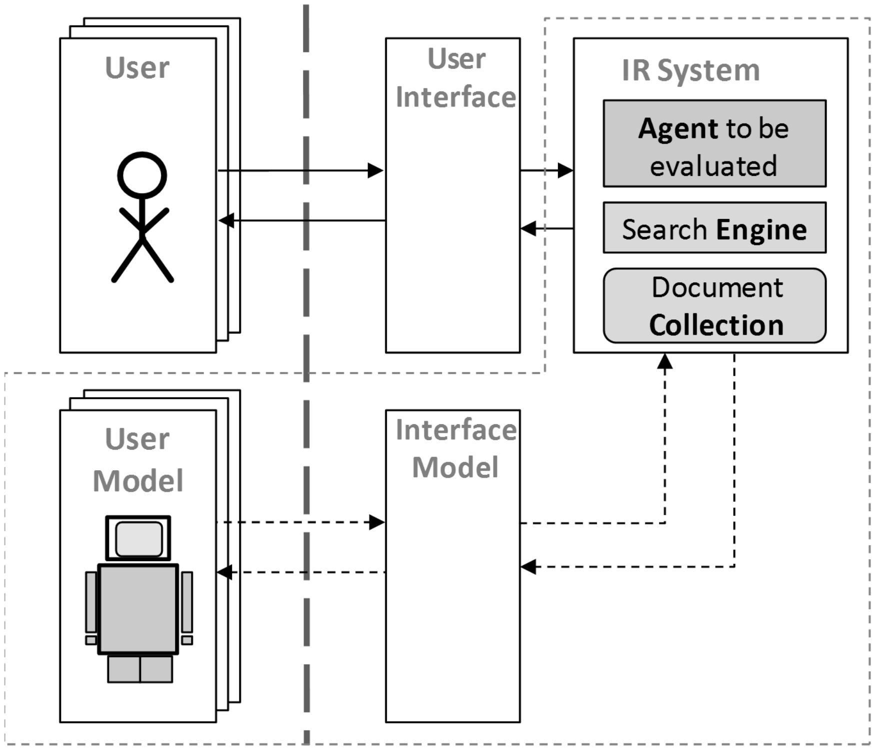



**Figure 4.1:** Schematic depiction of elements within an IR simulation **[Azzopardi et al. 2011]**

As part of the research conducted for and presented in this dissertation, the simulator for CIR tasks was developed in a series of Bachelor-theses conducted by students of the University of Hagen and mentored by the author of this dissertation. Contributions to the simulation software have been made by the students:

- Robert Dronsgalla, "*Testkollektionen basierte Evaluation in EzDL*", 2014
- Christian Steiner, "*Erweiterung eines Information-Retrieval-Evaluationswerkzeugs um iterative Suchsitzungen*", 2014
- Jan Lagerpunsch, "*Erweiterung eines Information-Retrieval-Evaluationswerkzeugs um kollaborative Suchsitzungen*", 2015

## 4.1 Evaluation Frameworks for Information Retrieval

IR researchers and practitioners have long implemented specialized toolkits and algorithms for document indexing and retrieval using dedicated evaluation frameworks. Such frameworks allow for rapid prototyping and gathering experimental data for thorough analysis of models and comparison with standardized baselines. Two of the most prominent ones are the following:

1. **Terrier** stands for **TE**rabyte **R**etr**IE**ve**R**, and is a modular and scalable framework for the rapid development of large-scale IR applications. Terrier is written in Java and is developed at the School of Computing Science at the University of Glasgow. It provides indexing and retrieval functionalities, as well as a number weighting models [Plachouras et al. 2004]. Terrier is provided including configurations that allow for indexing typical TREC test collections[6].

2. The **Lemur** software and datasets are another toolkit widely used in scientific research, see for example [Zuccon et al. 2011]. The Lemur toolkit includes the Indri search engine and data resources, such as the ClueWeb09 dataset, that support research and development of IR and text mining software. The Lemur Project was begun by the Center for Intelligent Information Retrieval at the University of Massachusetts and the Language Technologies Institute (LTI) at Carnegie Mellon University[7].

Whereas the above mentioned evaluation toolkits were mainly designed for system-oriented IR evaluation, i.e., they provide means to adequately compare the relative effectiveness of two retrieval strategies or algorithms, the framework **ezDL** [Beckers et al. 2012] has been developed to provide means for an user-oriented evaluation. For example, Kriewel and Fuhr [Kriewel and Fuhr 2010] utilized ezDL for user experiments to evaluate their adaptive search suggestion technique. ezDL is the continuation of the Daffodil [Klas et al. 2008] project and implements Meta-search in digital libraries and strategic support for users [Beckers et al. 2012]. It connects to remote search services (e.g. digital libraries) using wrapper agents and allows for aggregation of search results from several external search services[8].

---

[6] Please visit http://ir.dcs.gla.ac.uk/terrier for more details
[7] Please visit http://www.lemurproject.org for more details
[8] Please visit http://www.ezdl.de for more details

## 4.2 YaRS: Yet another Retrieval Simulator

Because ezDL (and its predecessor implementation Daffodil) has been used for evaluation in research projects related to this dissertation, such as [Landwich 2012], I decided to enhance the ezDL framework by a component dedicated to test-collection based evaluation. The Bachelor-thesis by Robert Dronsgalla implemented such functionalities as an additional ezDL agent. Because this Bachelor-thesis was only a starting point, it covered ad-hoc IR tasks of single users only. This work has been continued by the student Christian Steiner who enhanced the developed component towards evaluation of multi-query sessions based on [Keskustalo et al. 2009]. Finally, the student Jan Lagerpusch furthermore enhanced the developed component towards evaluation of collaborative, i.e. multi-user, sessions based on [Joho et al. 2009].

During the preparations of latter two theses, however, it has been decided to provide the simulation software as a stand-alone program instead of an integrated ezDL agent. This decision was made because of the high system complexity of ezDL and the resulting maintenance efforts. This stand-alone simulation software has been labeled **YaRS**, which stands for **Y**et **a**nother **R**etrieval **S**imulator, and is available to public in a CodePlex[9] code archive, to allow for reproduction and verification of results presented in the next chapter.

The following sections briefly introduce the software architecture of the simulator as well as the design decisions made.

### 4.2.1 Chosen Technology

#### 4.2.1.1 Programming Language

Java is a general-purpose computer programming language and specifically designed as platform independent language with integrated automatic memory management. Java code is typically compiled to byte-code that can run on any Java Virtual machine (JVM) regardless of the underlying computer architecture. A garbage-collector is responsible for recovering the memory once objects are no longer in use.

[9] To be found at: http://yars.codeplex.com

#### 4.2.1.2 Text Retrieval Library

Apache Lucene is a text search engine library written entirely in Java. It provides indexing capabilities including stemming and multi-valued fields. Lucene also includes several ranking models, such as the Vector Space Model and Okapi BM25.

#### 4.2.1.3 Linear Programming

As solver for ILP s, I used the numeric solver `lp_solve`[10] which is available as binary program for several platforms, including Windows x64. To be able to use the solver within Java programs, I used the Java wrapper `javailp`[11] which provides a set of APIs that allow for consuming the functionality of `lp_solve`. Please refer to the online documentation of the mentioned tools and libraries for installation instructions. An example program demonstrating how to solve an ILP using these tools and libraries is included in the mentioned CodePlex code archive.

### 4.2.2 User Model and Interface Model

The interface model definition aimed for the *"smallest common denominator approach"*, that is, the definition of basic user-interface elements of the IR system that are assumed to hold for most IR system interfaces. Implicitly, this also defined the user model, i.e., the sequences of actions performed by the simulated users.

More precisely, the IR interface allows a simulated user to enter and ***issue*** a query. As response, the IR interface provides a linearly ranked list of documents with a certain, user-specific length. The simulated user ***examines*** (inspects and assesses) the documents one by one (in linear order) starting at the top ranked document. The IR system allows the simulated user to provide a relevance assessment for each document. In case of real users, the latter step would correspond to bookmarking a Web-site or storing a document on the local disk.

Each simulated user processes the whole list of retrieved documents. Each relevant document is accumulated in a session-result. Please note that this user model corresponds to the assumptions of the user behavior made in section 3.3.4.

[10] To be found at http://lpsolve.sourceforge.net/5.5/
[11] To be found at http://javailp.sourceforge.net/

### 4.2.3 Software Architecture of YaRS

The diagram in Figure 4.2 concretizes the elements within an IR simulation of the general schematic depiction in Figure 4.1. The simulator roughly consists of three main software components for which a schematic depiction is provided in Figure 4.2 and explained below. These components are ***Engine***, ***Agent*** and ***Client***, and are described in more detail below. Note that the three main components correspond to Java packages; sub-components correspond to Java classes of the corresponding packages.

I used the graphical notation defined by the Fundamental Modeling Concepts[12] (FMC), an industry driven consortium that provides a framework for the comprehensive description of complex software systems. This framework is based on a precise terminology and graphical notation. I used a Block Diagram[13] to depict the compositional structure of the simulation software. The diagram in Figure 4.2 is composed of agents (active components that access adjacent passive system components like storages), storages (passive system components used to store data) and channels (edges indicating access of an active component to a passive component).

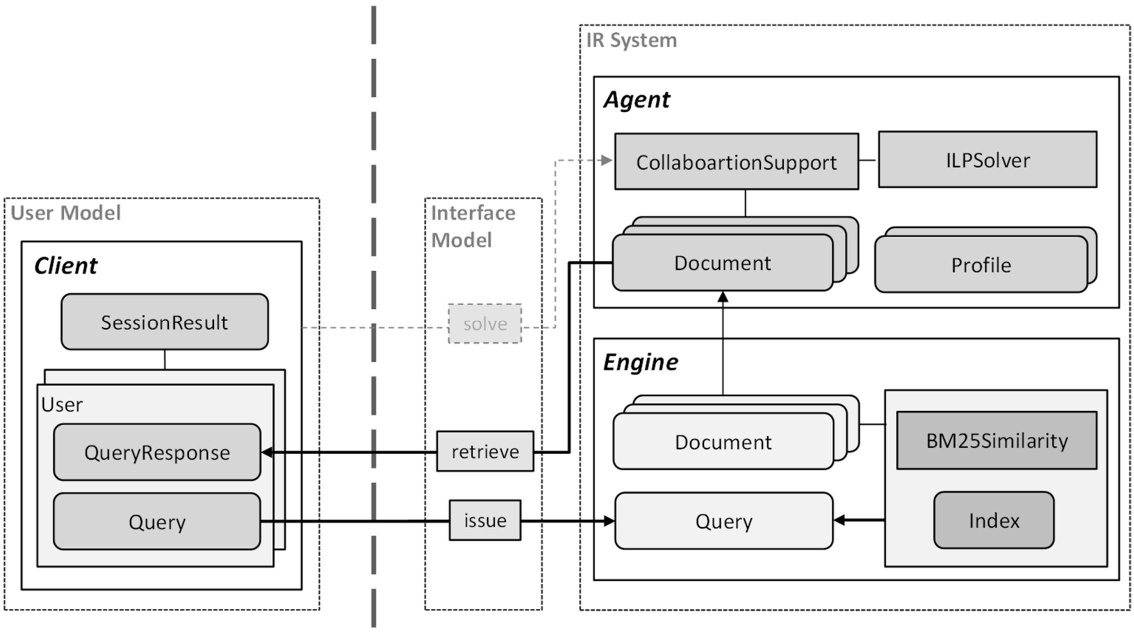



**Figure 4.2:** Architectural depiction of the simulation tool and its main components

[12] Please visit http://www.fmc-modeling.org for more information
[13] Please visit http://www.fmc-modeling.org/notation_reference for the notation reference

1. **Engine**. This component represents the search engine and executes given queries against an indexed document corpus and returns a set of documents as response. It is implemented in the Java package `org.hagen.yars.engine.`

   This component is composed of elements of the Apache Lucene libraries and provides standard search engine functionality based on BM25 similarity. Entities, such as documents and queries, are implemented using class definitions provided by the Apache Lucene package.

2. **Agent**. This represents the collaboration support component and implements the optimum criterion for collaborative search developed in section 3.4 as a dedicated software functionality (or service). It is implemented by the Java package `org.hagen.yars.agent.`

   A detailed description of the implementation of this component can be found in the section 4.3.

3. **Client.** This component implements the process of an experiment run, that is, several simulated users select and issue a query, retrieve results, and store search results identified as relevant in an accumulated session result. It is implemented by the Java package `org.hagen.yars.client.`

   This component consists of the following sub-components:

   a) Simulated users are implemented by the sub-component *User*. Each simulated user is associated with a query response (*QueryResponse*), i.e., documents returned by the collaboration support component in response to the query processed by the search engine component.

   b) The *SessionResult* represents the accumulated set of documents found by all simulated users (i.e. the team) involved in the experiment. This component also provides means for calculating evaluation measures.

Moreover, another component (**Collection**) which is not depicted in Figure 4.2 provides means for reading and parsing an IR test-collection. Please note that the process of index creation (which requires the means of the Collection component) is performed in an off-line step before the simulation runs and therefore not depicted in the figures.

### 4.2.4 Data Model of YaRS

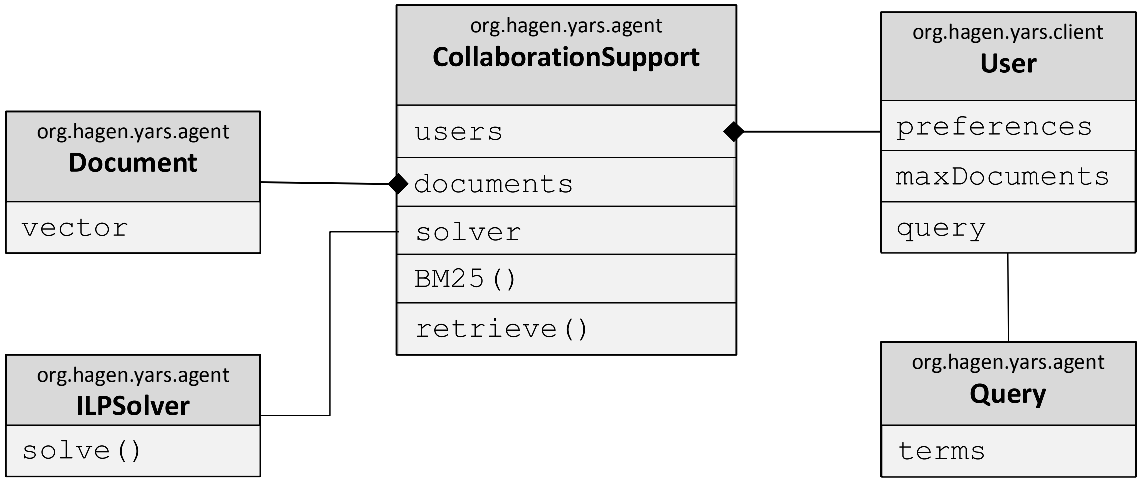



**Figure 4.3:** Illustration of data model of YaRS using a class diagram of involved entities

Figure 4.3 schematically depicts the abstract data structures used for implementation of YaRS as well as associations among them using a class diagram. Java class `org.hagen.yars.Document` implements a corpus document, where attribute `vector` is an array that represents the document's content as term vector weighted using the Term Frequency–Inverse Document Frequency (TF-IDF) value. This is realized using a `SortedMap<String, Double>`.

A query is implemented by Java class `org.hagen.yars.agent.Query`, where attribute `terms` is an array of strings that represents the query terms. Users are implemented by Java class `org.hagen.yars.client.User`. This class includes a weighted keyword user profile (attribute `preferences`), realized as `TreeMap<String, Double>`.

Access to the Activity Suggestion implementation (see section 4.3) is provided via Java class `org.hagen.yars.agent.CollaborationSupport`. A client aiming to utilize this service will need to provide instances of `users` and `documents` via corresponding setter-methods (not illustrated in Figure 4.3). Based on the provided data, method call `solve()` internally creates the results by utilizing the `ILPSolver` instance stored in attribute `solver`. Result sets (arrays of document instances) can be requested by the client using method call `retrieve()` for each simulated user. The following figure (Figure 4.4) illustrates this call sequence.

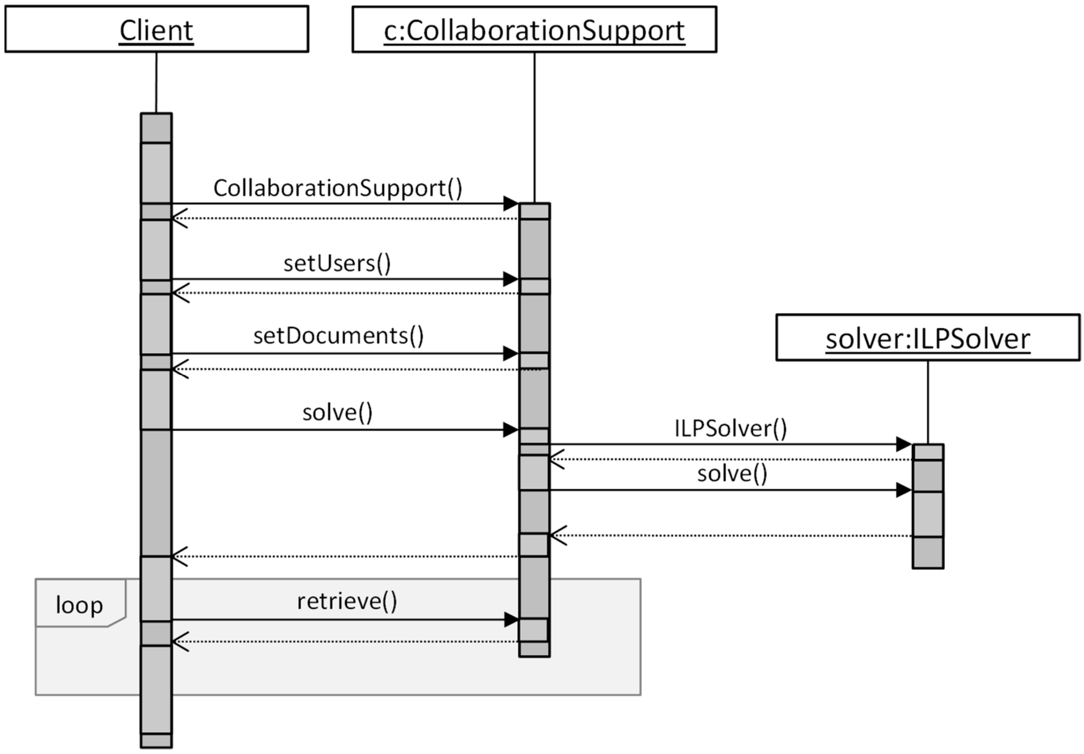



**Figure 4.4:** Sequence diagram of the general call sequence of a simulation run in YaRS

## 4.3 Implementation of Activity Suggestions

This section outlines how the formal criterion for optimum collaboration, i.e., Activity Suggestions, as developed in chapter 3 is implemented within YaRS. Mainly, the implementation of YaRS re-uses means provided by several (freely available) Open Source software libraries. This covers the estimation of probabilities of relevance (4.3.1) as well as linear programming (section 4.3.2). Due to the fact that the implementation of Activity Suggestions is based on functionality of existing Open Source libraries, this section also demonstrates how easy it is to implement Activity Suggestions and integrate it into an IR system.

### 4.3.1 Probability Estimation

Major element of Activity Suggestions developed in section 3.4 is the estimation of probability of relevance of a document given a query issued by a user. Typically, such estimations are performed in every modern search engine employing a probabilistic retrieval model. Most prominent procedure of doing this is employing the BM25 model [Robertson et al. 2004] which calculates scores that are rank-proportional to the probabilities of relevance. In my implementation, I make also use of BM25. This calculation assumes a document represented as weighted term-vector and a query represented as "*bag-of-words*". However, in its latest version, the BM25 ranking formula has been extended by Robertson et al. to deal with weighted terms. As shown in [Bouhini et al. 2014], this can be used to incorporate a weighted keyword profile into the ranking function to personalize and refine the search result. The following code snipped (Listing 4.1) depicts the method's signature for computing a score according to the BM25 model, as it is found in the YaRS project. The implementation of this method is taken from Apache Lucene, specifically from Java class `BM25Similarity()` of the `search.similarity` package.

```
public double BM25(  Document document,
                     String[] query,
                     TreeMap<String, Double> preferences );
```

**Listing 4.1:** Signature of method `BM25` of class `CollaborationSupport`

### 4.3.2 Linear Programming

The Java project `javailp` provides a simple interface that allows for defining and solving an ILP which represents my optimum criterion for collaborative search session (see section 3.4).

At implementation level, the optimum criterion is represented by a so called problem which is an instance of `javailp.Problem`. A problem consist of an array of vectors of numerical values plus an additional condition associated with each of these vectors. The first one of these vectors describes the objective, i.e., the sum to be maximized. The other vectors describe the constraints of the optimum criterion. Each of these vectors is implemented using Java class `javailp.Linear.`

The following code snipped (Listing 4.2) illustrates the usage of the Java classes `javailp.Linear` and `javailp.Problem.` Both classes are used for defining the objective of the developed optimum criterion, i.e. $\max \sum_i^N \sum_j^M s_{i,j}\rho_{i,j}$. The objective is a sum of $N \times M$ elements and has the following structure:

$$\begin{aligned}\max\big(&s_{1,1}\rho_{1,1} + s_{1,2}\rho_{1,2} + \dots + s_{1,M}\rho_{1,M} + \\ &s_{2,1}\rho_{2,1} + s_{2,2}\rho_{2,2} + \dots + s_{2,M}\rho_{2,M} + \\ &\dots \\ &s_{N,1}\rho_{2,1} + s_{N,2}\rho_{N,2} + \dots + s_{N,M}\rho_{N,M}\big)\end{aligned} \tag{5.1}$$

In the code snipped below (Listing 4.2), the constant `OptType.MAX` indicates a maximization problem. A string concatenated of `user.ID` and `document.ID` serves as unique identifier of the vector elements.

```
public void solve(Document documents, User users) {

  Problem problem = new Problem();
  Linear vector = new Linear();

  for( Document document : documents ) {
    for( User user : users ) {
      vector.add( BM25( document, user.query ),
                  user.ID + "" + document.ID );
    }
  }
  problem.setObjective(objective, OptType.MAX);
```

**Listing 4.2:** Code snipped of method `solve` of class `ILPSolver`

The first constraint of the optimum criterion, i.e. $\sum_i^N s_{i,j} \leq 1, \forall j$, is basically a set of *M* inequalities of the following form:

$$\begin{aligned}&s_{1,1} + s_{2,1} + \dots + s_{N,1} \leq 1 \\ &s_{1,2} + s_{2,2} + \dots + s_{N,2} \leq 1 \\ &\dots \\ &s_{1,M} + s_{2,2} + \dots + s_{N,M} \leq 1\end{aligned} \tag{5.2}$$

The following code snipped (Listing 4.3) depicts the definition of these inequalities using the interfaces of package `javailp`.

```
  for( Document document : documents ) {
    vector = new Linear();
    for( User user : users ) {
      vector.add( 1, user.ID + "" + document.ID );
    }
    problem.add( vector, "<=", 1 );
  }
```

**Listing 4.3:** Continued code snipped of method `solve` of class `ILPSolver`

Similarly, the second constraint, i.e. $\sum_j^M s_{i,j} \leq K_i, \forall i$, can be defined using the same means. It consists of $N$ inequalities of the following form.

$$\begin{aligned} &s_{0,1} + s_{0,1} + \cdots + s_{0,M} \leq K_0 \\ &s_{1,1} + s_{1,1} + \cdots + s_{1,M} \leq K_1 \\ &\ldots \\ &s_{N,1} + s_{N,2} + \cdots + s_{N,M} \leq K_N \end{aligned} \tag{5.3}$$

However, the condition on the right hand side of the inequalities are the user specific assessment capacities which can be accessed by method `user.getCapacity()`. The following code snipped (Listing 4.4) depicts the definition of these inequalities using the interfaces of package `javailp`.

```
  for( User user : users ) {
    vector = new Linear();
    for( Document document : documents ) {
      vector.add( 1, user.ID + "" + document.ID );
    problem.add( vector, "<=", user.getCapacity() );
  }

  Solver solver = factory.get();
  result = solver.solve(problem);
}
```

**Listing 4.4:** Remaining code snipped of method `solve` of class `ILPSolver`

Finally, using method call `result.getPrimalValue(user.ID + "" + document.ID)` (not part of the code snipped above), single elements of the ILP's solution can be obtained. So, by specifying a particular user and document, each specified via the corresponding identifier (user ID and document ID), the return value of this method indicates whether the document is suggested to the user (integer value of 1) or not (integer value of 0).

## 4.4 Summary

This chapter demonstrated the practicability of the developed formal criterion by a prototypical implementation of Activity Suggestions as a component within an IR system. This chapter also demonstrated that the implementation could be realized by the re-use and combination of functionalities provided by various Open Source software libraries. For example, an existing solver for ILPs could be integrated easily using a Java wrapper API. Both, the ILP solver as well as the wrapper API are Open Source projects and, thus, freely available.

The next chapter employs the prototypical implementation for a quantitative evaluation which exemplified the application of Activity Suggestions by means of search result division of query responses among team members.

# 5 Evaluating CIR Support

This chapter presents a thorough evaluation based on a computer simulation of collaborative search tasks in two professional domains. This simulation employs simulator tool and the prototypical implementation of Activity Suggestions outlined in the previous chapter. Based on the evaluation methodology described in section 5.1, the approach developed in this dissertation is compared with several baselines. The effects of a changing team size are studied as well as the effects of a changing number of documents examined by each team member. Evaluation results and significance tests are presented in section 5.2.

## 5.1 Evaluation Methodology

Simulations of users' interactions have been used extensively in IR research to evaluate IR systems that support individual users as well as collaborative teams. This covers, e.g., implicit relevance feedback [White et al. 2005], explicit relevance feedback [Järvelin 2009], query expansion for single users [Ruthven 2003], and collaborative users [Hust 2005]. Also, evaluation of collaborative search systems with backend mediated collaboration have been conducted by means of simulations: For example, Pickens et al. showed how their algorithm could achieve an effective collaboration by way of simulation [Jeremy Pickens et al. 2008], Shah et al. demonstrated how search processes that were virtually combined could result in achieving results that are both relevant and diverse [Chirag Shah et al. 2010]. Foley and Smeaton as well as Soulier et al. demonstrated the effectiveness of their models by simulating users searching together synchronously based on interaction logs of individual users from the TREC interactive track experiments [Foley and Smeaton 2010] [Soulier et al. 2013].

According to White et al. [White et al. 2005], the benefits of simulations are:

1. They are less costly and time consuming, compared to real user experiments.
2. They allow for the evaluation of IR techniques in many different retrieval scenarios.
3. The experimental setup can be controlled by the system designer.

In this dissertation, I also apply simulation as experimental methodology to explore different search result division strategies that could be employed by a collaborating team. I chose this methodology because it provides adequate means for comparing the relative effectiveness of different strategies or algorithms [Voorhees 2000].

In this section, I detail and justify the simulated collaboration that was employed to generate the data used in the analysis (section 5.2.3). I describe the materials used for my experiments (section 5.1.1) which cover the utilized datasets, employed software tools, and applied measures. Afterwards, I present the method used to simulate the collaborative search processes (section 5.1.3) which covers the user model (querying and assessment behavior of simulated users) and baseline procedures.

### 5.1.1 Materials

Experiments were conducted using two IR test-collections covering two domains of interest: the intellectual property domain and medical domain.

Both of the utilized test-collections have either been used in TREC experiments [Robertson and Hull 2000] or CLEF experiments [Roda et al. 2010]. The Text REtrieval Conference (TREC) was established in 1992 and co-sponsored by the National Institute of Standards and Technology (NIST) and the Intelligence Advanced Research Projects Activity (IARPA) of the United States of America. It is an on-going series of workshops focusing on a list of different IR research areas (also called tracks) each evaluating different aspects of the IR process. A related series of workshops is the Cross-Language Evaluation Forum (CLEF) which also aims to maintain an underlying framework for evaluating IR systems and creating repositories of data for researchers to have comparable standards.

#### 5.1.1.1 The OHSUMED Collection

I selected OHSUMED test-collection [Hersh et al. 1994] for the simulation of a retrieval task upon the request of disease information on a medical literature corpus. The OHSUMED document corpus was used, for example, in the TREC-9 Filtering Track [Robertson and Hull 2000].

The OHSUMED corpus is composed of 348.566 MEDLINE documents from 270 journals published between 1987 and 1991. Each document contained in this collection has several attributes from which I considered the text-attributes title and abstract for indexing. An example document is presented in Table 5.1.

This test-collection includes 106 topics. Table 5.2 presents one example from the topics including all of its attributes. OHSUMED contains relevance assessments manually annotated using three relevance levels (“*definitely relevant*”, “*possibly relevant*”, and “*not relevant*”). I considered both, “*definitely*” and “*possibly relevant*”, as relevant. The relevance assessments are contained in a separate text file of this test-collection which represents a simple mapping between topic identifier and document identifiers, i.e., for each topic identifier, a list of document identifiers is provided.

**Table 5.1:** Example of OHSUMED document with selected attributes

| | |
|---|---|
| **MEDLINE identifier** | 91000003 |
| **Title** | Treatment of Fournier's gangrene with adjunctive hyperbaric oxygen therapy. |
| **Authors** | Lucca M; Unger HD; Devenny AM. |
| **Abstract** | Fournier's gangrene is a devastating infection and often is associated with a high morbidity and mortality. Surgical debridement and antibiotics are the cornerstones of therapy. This case describes the use of hyperbaric oxygen as an adjunct in the treatment of Fournier's gangrene. |
| **Source (and publication type)** | Am J Emerg Med 9101; 8(5):385-7 (Journal Article) |

**Table 5.2:** Example of OHSUMED topic with all attributes

| | |
|---|---|
| **Identifier** | 1 |
| **Patient information** | 60 year old menopausal woman without hormone replacement therapy |
| **information need** | Are there adverse effects on lipids when progesterone is given with estrogen replacement therapy |

### 5.1.1.2 The CLEF-IP Collection

The task for the intellectual property domain is a patentability search which aims to find patents that constitute prior art and may conflict with a new patent [Joho et al. 2010]. As patent corpus, I used the CLEF-IP 2009 corpus [Roda et al. 2010]. This test-collection was used in the CLEF-IP track that was launched in 2009 to investigate IR techniques for patent retrieval.

The CLEF-IP corpus consisted of 1.958.955 patent documents pertaining to 1,022,388 patents with publication date between 1985 and 2000. In general, one patent (identified by a unique patent number) corresponds to several patent documents generated at different stages of the patent's life-cycle. Each patent document contained in this collection has several attributes from which I only chose the text attributes title and the abstract for indexing. An example patent document is presented in Table 5.3.

I indexed the documents according to the CLEF-IP 2009 track guideline[14], that is, combine the patent documents in a *"virtual"* publication by taking each field from the latest publication and index this *"virtual"* patent. Please note that I only considered text attributes of patent documents in English for indexing.

The CLEF-IP 2009 test-collection also contains topic definitions and relevance assessments. A topic names the patent to which prior art is to be identified. The relevance assessments list patents constituting the prior art. Notably is that relevance is measured on patent level not on patent document level.

I decided to use these two test-collections, since both are freely available which allows for easier reproduction of my experimental results. Also, simulation of professional search in the medical domain as well as in the intellectual property domain have recently been conducted using the OHSUMED collection and a patent corpus [Kim et al. 2011], although a different patent collection (an American one) was used for the patent retrieval task by Kim et al.

[14] To be found at http://ifs.tuwien.ac.at/~CLEF-ip/

**Table 5.3:** Example of CLEF-IP patent document with selected attributes

| | |
|---|---|
| **Document identifier** | EP-0274375-A2 |
| **Title** | Magnetic track brake actuator. |
| **Inventors** | Rocholl, Hans |
| **Assignee** | Bergische Stahl-Industrie<br>Papenbergerstrasse 38<br>D-42859 Remscheid |
| **Agent** | Jung, Hermann L., Dipl.-Chem.<br>Patentanwalt Postfach 17 28<br>61287 Bad Homburg v.d.H. |
| **Abstract** | In the magnetic track brake actuator, the magnetic brake can be lowered and lifted, a damping ring (19), made of an elastic or elastomeric material, being arranged between the cylinder lid (12) and the attachment flange (14). The damping ring has an equal or greater damping capacity than the bearing ring (15, 16) used and the bearing ring or bearing rings (15, 16) have the same or greater hardness than the damping ring (19). |

From the topics of the test-collections, I selected only those topics that had at least 20 relevant documents assigned. This cut-off ensured that there were enough relevant documents for examining collaboration and search result division among co-searchers, i.e., if we only had a few of relevant documents per topic it is likely that only one user could have accomplished the search task (find a satisfying amount of relevant documents) which would not be particularly interesting. This left 68 topics out of the OHSUMED collection and 144 out of the CLEF-IP collection. Statistics about the considered topics are summarized in Table 5.4.

**Table 5.4:** Statistics about relevant documents of remaining topics

| **Test-Collection** | **No. of selected topics** | **No. of relevant document per selected topic** | | | |
|---|---|---|---|---|---|
| | | **Minimum** | **Maximum** | **Mean** | **Median** |
| **OHSUMED** | 68 | 20 | 149 | 52.22 | 43 |
| **CLEF-IP** | 144 | 20 | 50 | 27.53 | 26 |

### 5.1.2 Measures and Tools

I chose the ***Group Recall*** [Baeza-Yates and Pino 1997] as measure of retrieval effectiveness because my research interest in this dissertation covers recall-oriented search tasks. This measure has been introduced as equation 2.2 in section 2.6.4. It builds an overall (relevant) result-set by merging all relevant documents found by the individual team members, i.e., *RelevantCoverage*(*T*). The merging procedure is the union set operation. The recall is computed by dividing the overall number of relevant documents found by team members by the number of documents annotated as relevant for a test-collection topic, i.e., $U_R$.

My simulator is a Java-based application that uses Apache Lucene 4.9 for indexing and searching. As numeric solver for ILPs, I used `lp_solve`[15] with the Java wrapper `javailp`[16]. Documents have been indexed using Porter's stemming and a standard stop-word list for English text. A detailed description of the simulation software can be found in chapter 4.

### 5.1.3 Simulation Methods

#### 5.1.3.1 Simulation of Collaboration

In the CIR domain, there are no official test-collections nor baselines to be used for evaluation and comparison of techniques. Therefore, to obtain results that allow for comparison with prior experiments in the field of CIR, I chose a simulation procedure used for evaluation in [Chirag Shah et al. 2010] and similarly in [Joho et al. 2009]. It consisted of the following steps:

1) Each simulated user issued a query.

2) Documents of all query responses were merged into a shared result-set using the CombSUM algorithm that combines the scores of all users' queries [Chirag Shah et al. 2010].

3) From this shared result-set, each simulated user was provided with a result-page for assessment consisting of $K_i$ documents. In my experiments, result-pages were extracted from the shared result-set either by applying one of the baseline procedures (see section 5.1.3.5) or by applying my optimum criterion (see section 3.4.1).

Figure 5.1 schematically depicts the employed simulation procedure. In accordance to the process model of CIR support of section 3.2.2, besides an IR system, the simulated environment contains a CIR support component. We see the flow of interactions exemplified by two simulated users. Each of them passes through the stages of the Information Dialog (see section 3.3.4). Different methods of result-page extraction

[15] To be found at http://lpsolve.sourceforge.net/5.5/
[16] To be found at http://javailp.sourceforge.net/

are element of investigation in the experiments. The CIR support component generates result-pages according to the methods described in section 5.1.3.5. We see that:

1. Each simulated user issued a query (stage Access) against the corpus $D$ of documents contained in the information source. As response to this query, the simulated user obtains search result (see definition 4.1) that is stored in the shared result set $D^{(l)}$ (see definition 4.5).

2. From this shared set $D^{(l)}$, each simulated user requests a result-page (stage Orientation). This result-page $D_i^{(l)}$ represents a sub-set of $D^{(l)}$ (see definition 4.8). Several methods of creating this result-page (including the developed optimum criterion) are studied in the experiments.

3. Each document of the result-page is judge with respect to relevance (stage Assessment). Relevant documents of all users are accumulated in the set *RelCov* (see definition 4.7) which represents the assessment outcome of the whole team.

Please note that I employed this procedure for each topic of both test-collections and for all considered result-page sizes, leading (in sum) to 1696 simulated session of collaboration.

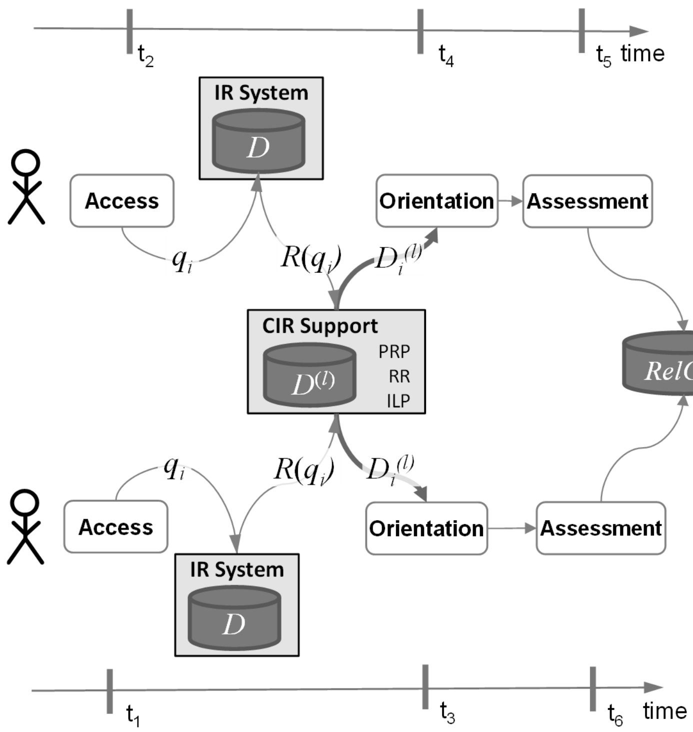



**Figure 5.1:** Depiction of interaction-flow of two simulated users.

5.1.3.2 Result Page-Size

Depending in the search task type, professional searchers typically assess (on average) 100 documents per query (see [Joho et al. 2010], table 2). In my experiments, I varied this number as I was interested in its influence on collaborative retrieval effectiveness. Also, I wanted to cover the typical assessment behavior identified for several professional search tasks. For example, professionals examine (on average) 50 documents when engaged in a patentability search.

I considered the following sizes of result-pages that simulated users examine on average $K \in \{20, 50, 80, 120, 160, 200, 260, 320\}$. I selected these sizes to cover the top ranks and first few pages of search results. Also, I consider significantly deeper depths (or multiple pages of search results) because even it might be unlikely that users would, on average, examine hundreds of documents per query, it is of interest to see whether this strategy is better or worse than other strategies.

5.1.3.3 Query Construction

Recent developments in formal models for simulating user querying behavior allows generating queries which obtain performance similar to the performance of actual user queries [Azzopardi et al. 2007] [Azzopardi 2009]. I used the query generation process examined by Azzopardi who modeled a user who selects terms from an imagined ideal relevant document:

$$P(t|query) = (1 - \lambda)P(t|topic) + \lambda P(t) \tag{6.1}$$

Sometimes, chosen terms will be on topic, $P(t|topic)$, while other times terms will be off topic, $P(t)$. The distribution $P(t|topic)$ describes the occurrences of terms in the ideal relevant document and relates to the users' background knowledge. For estimating $P(t|topic)$, I used the strategy called "*Frequent*" [Azzopardi 2009] which assumed that users are likely to select terms that stand out in some way so that more frequent terms are more likely to be used as query terms. In [Azzopardi et al. 2007], it was shown queries created using this strategy (called "*Popular*" in [Azzopardi et al. 2007]) were similar to real users' queries and also delivered performance that was most like that obtained from real queries. Because I needed to model different users, I decided to vary the $\lambda$ parameter slightly and, thus, selected this parameter randomly from the interval $\lambda \in \{0.1; 0.3\}$. This also reflects the amount of "*noise*" observed in real queries [Azzopardi et al. 2007]. The following equation represents the estimation of $P(t|topic)$ based on the strategy *Frequent*, where $U_R$ denotes the set of relevant documents for the topic, $n(t,d)$ is the number of times a term $t$ appears in a document $d$, and $n(d)$ is the number of terms in $d$.

$$P(t|topic) = \frac{\sum_{d \in U_R} n(t,d)}{\sum_{d' \in U_R} n(d')} \tag{6.2}$$

For each topic of the test-collection, this allowed me to generate weighted term vectors $w(U_R,t)$. I ranked $w(U_R,t)$ from highest to lowest. I decided to use three terms per query as this is a typical length for user queries [Arampatzis and Kamps 2008]. To obtain queries for simulated teams of up to six team members, I created tri-grams from the ranked vector $w(U_R,t)$. Meaning that the first team member used query term 1, 2 and 3, the second team member used query terms 3, 4 and 5, and so on. In this way, each user issued an own query. However, it is interesting to note that the queries produced by this generation method, far from being complete artificial, reflected the querying behavior observed in the analysis of search logs conducted by Foley and Smeaton. They found that queries issued by different people addressing the same information need most often have at least one term in common [Foley and Smeaton 2010]. For illustrative purposes, Table 5.5 exemplary presents generated queries for one selected topic of the OHSUMED collection. Six queries per topic are shown. The query terms have been stemmed using Porter's stemmer.

#### 5.1.3.4 Relevance Assessments of Simulated Users

For my experiments, I chose different team sizes of up to six team members. As in previous simulation work, e.g., [Joho et al. 2010] and [Chirag Shah et al. 2010], simulated users judged document relevance as the test-collection creators did, i.e., all relevant documents appearing in a result-page were counted as document found by a team member.

**Table 5.5:** Examples of generated queries based on an abstract of the OHSUMED collection

| | |
|---|---|
| **MEDLINE identifier** | `91000003` |
| **Title** | `Treatment of Fournier's gangrene with adjunctive hyperbaric oxygen therapy.` |
| **Generated Queries** | `estrogen cholesterol lipoprotein`<br>`estrogen increas women`<br>`hormon increas serum`<br>`hormon level therapy`<br>`effect group therapi`<br>`effect postmenopaus treatment` |

5.1.3.5 Baselines

I used the following two baselines for my experiments.

1) To extract a result-page from the shared result-set, for each user, I re-ranked the whole shared result-set according to the users' formalized query (estimated using BM25) and did a cut-off after $K_i$ documents. This approach simulated the case of team members employing search tools designed for individual usage, i.e., results are optimized towards an individual. I called this baseline **PRP**.

The obvious disadvantage is that the result-pages created for team members are likely to have an overlap. To avoid this overlap, I also used a baseline applied in [Chirag Shah et al. 2010] and [Joho et al. 2010].

2) The shared result-set is split using a Round Robin procedure. For example, if the team size was two, the first user got documents 1, 3, 5, etc., and the second user got documents 2, 4, 6, etc. Both distinct halves of the shared result-set (that are ranked according to the CombSUM score, see 5.1.3.1) are then re-ranked towards the team member's formalized query and cut-off after cut-off after $K_i$ documents. I called this baseline **RR**.

This baseline did both, avoiding redundancy and optimizing search results towards an individual. Please note that prior experiments also employed a *k*-means clustering baseline but the Round Robin procedure has been reported as the stronger baseline [Joho et al. 2010]. I therefore did not employ the *k*-means clustering baseline.

3) Finally, applying my optimum criterion is referred to as **ILP**.

## 5.2 Evaluation Results

This chapter presents the evaluation results that have been gathered using a computer simulation of collaborative search tasks with variations of several model parameters. I studied the effects of a changing team size as well as effects of a changing number of documents examined by each team member. The chapter addresses research question *RQ3* (see chapter 1).

Figures Figure 5.2 to Figure 5.7 present the results of my experiments for both test-collections used, OHSUMED and CLEF-IP. Diagrams in these figures depict the Group Recall as the number of documents examined per user change. The x-axis represents the number of documents inspected by each team member. The y-axis represents the Group Recall. Each diagram contains six curves that result from different team sizes. Each data point in the diagrams is an average of 68 samples and 144 samples, respectively (one per topic considered).

Moreover, I was interested in the gain that resulted from adding further team members to the search task. This is depicted in Figure 5.8 to Figure 5.10 for the CLEF-IP collection and in Figure 5.11 to Figure 5.13 for the OHSUMED collection. In these figures, the percentage gain of recall is plotted as the number of users change. More generally, these figures show the change of retrieval effectiveness which resulted from increasing the team size.

In the remaining sections of this chapter, the results are thoroughly discussed and analyzed. Furthermore, the significance of the gathered results is tested using a paired student's t-test along with a presentation of the corresponding statistics. This chapter closes with a discussion of the generalizability of the results and its limitations.

### 5.2.1 Retrieval Effectiveness

#### 5.2.1.1 CLEF-IP Collection

**Figure 5.2:** Group recall based on the **CLEF-IP** collection and baseline **PRP**

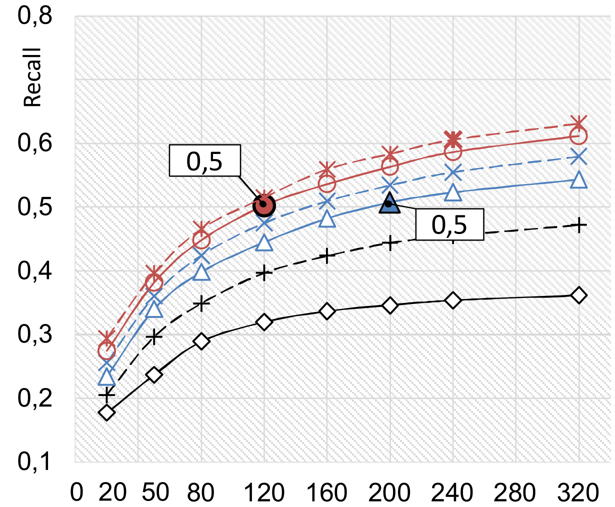



**Figure 5.3:** Group recall based on the **CLEF-IP** collection and baseline **RR**

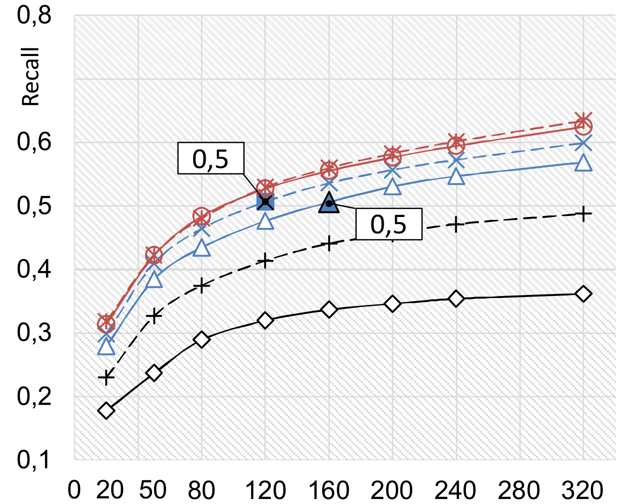
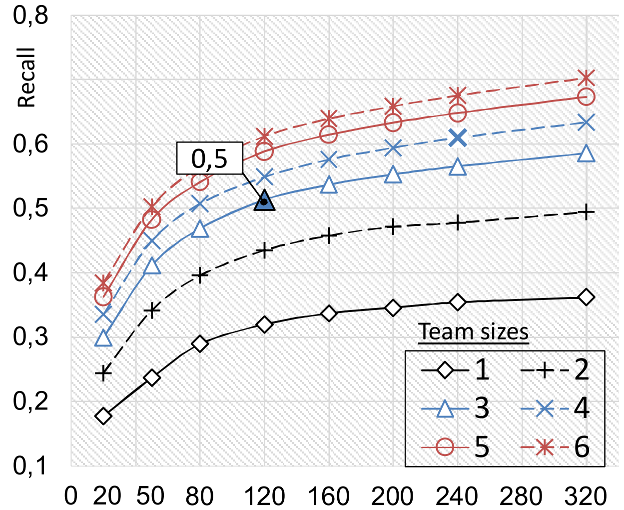



**Figure 5.4:** Group recall based on the **CLEF-IP** collection and **ILP** approach



### 5.2.1.2 OHSUMED Collection

**Figure 5.5:** Group recall based on the **OHSUMED** collection and baseline **PRP**

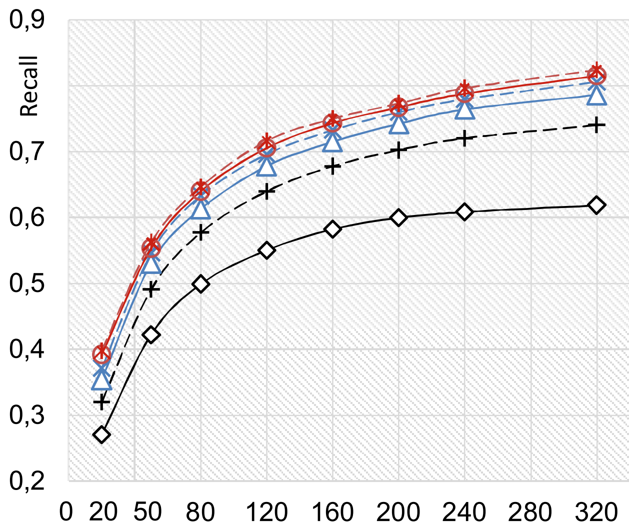



**Figure 5.6:** Group recall based on the **OHSUMED** collection and baseline **RR**

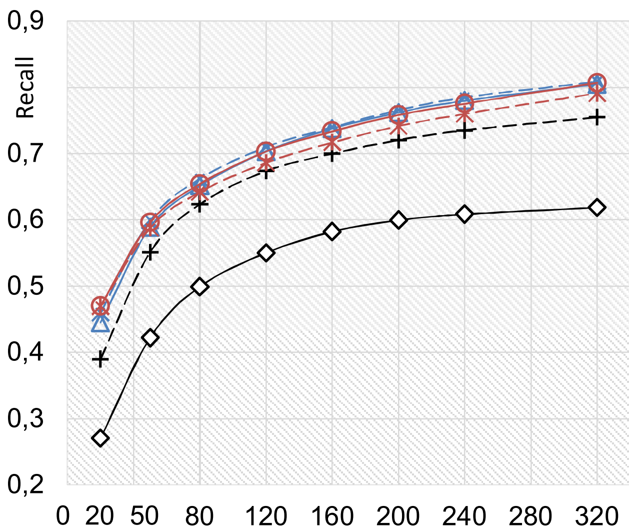
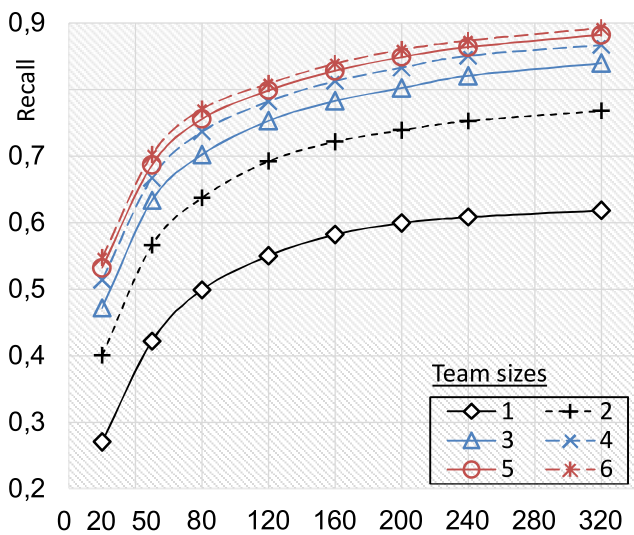



**Figure 5.7:** Group recall based on the **OHSUMED** collection and **ILP** approach



### 5.2.2 Percentage Gain of Recall

#### 5.2.2.1 CLEF-IP Collection

**Figure 5.8:** Plot of percentage gain based on the **CLEF-IP** collection and baseline **PRP**

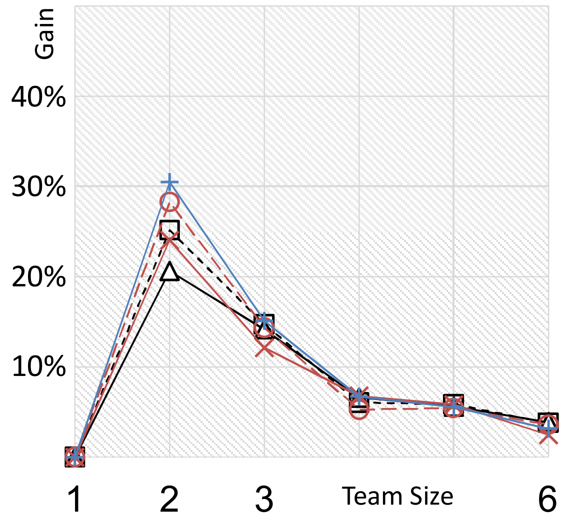



**Figure 5.9:** Plot of percentage gain based on the **CLEF-IP** collection and baseline **RR**

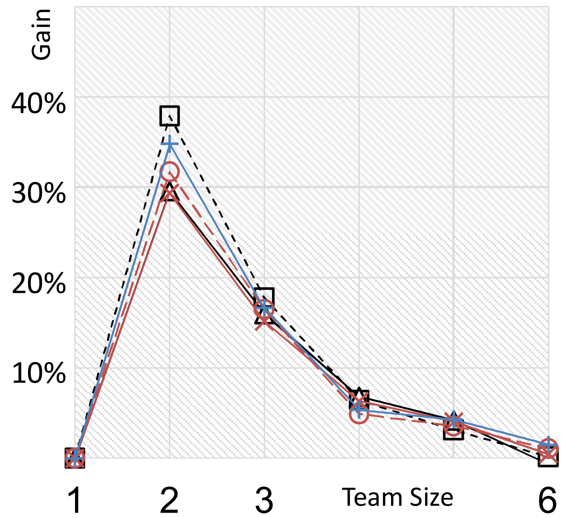
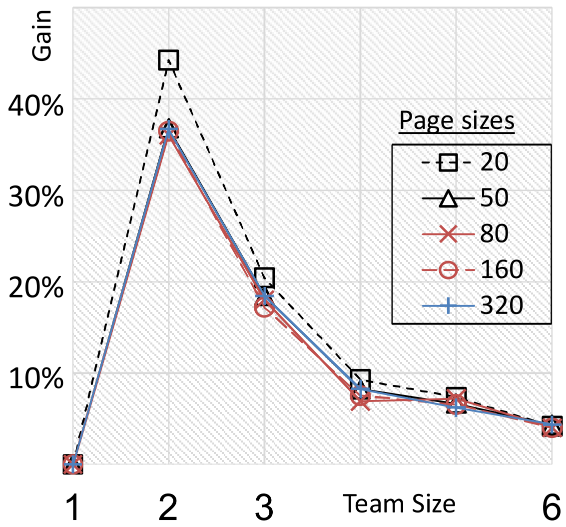



**Figure 5.10:** Plot of percentage gain based on the **CLEF-IP** collection and **ILP** approach



#### 5.2.2.2 OHSUMED Collection

**Figure 5.11:** Plot of percentage gain based on the **OHSUMED** collection and baseline **PRP**

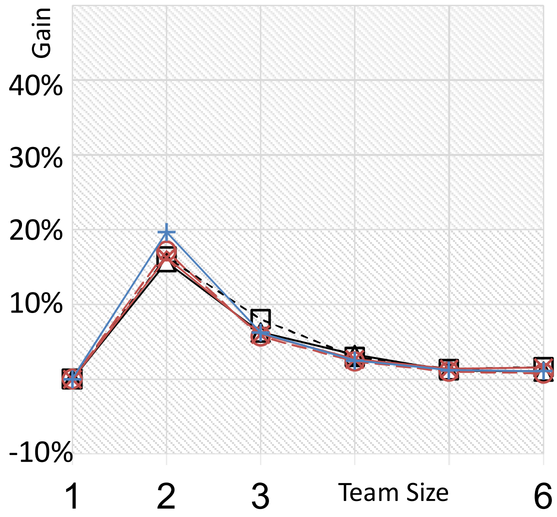



**Figure 5.12:** Plot of percentage gain based on the **OHSUMED** collection and baseline **RR**

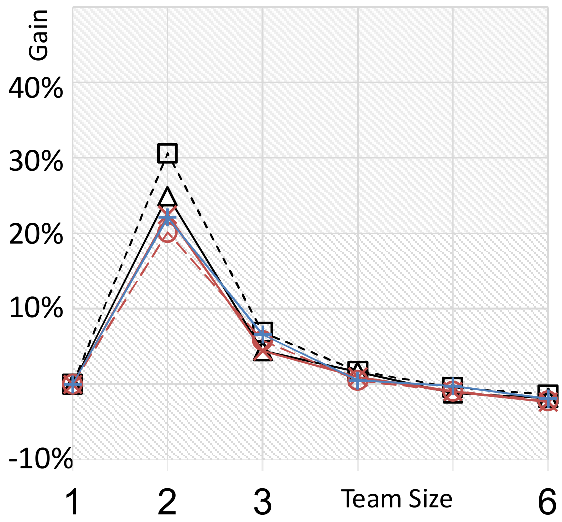
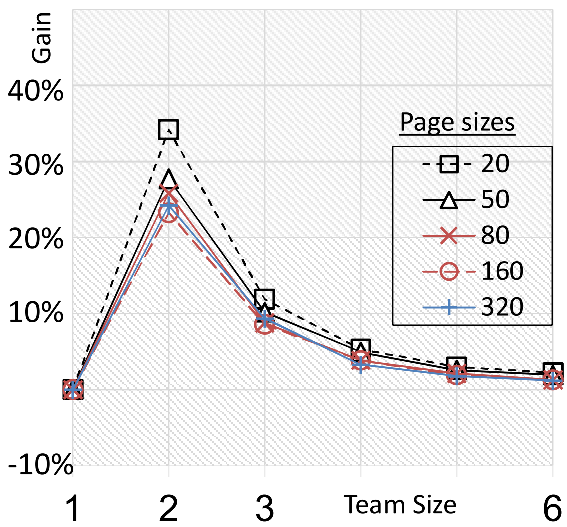



**Figure 5.13:** Plot of percentage gain based on the **OHSUMED** collection and **ILP** approach



### 5.2.3 Analysis

Figure 5.2 to Figure 5.4 summarize the results of my experiments based on the CLEF-IP collection. Figure 5.5 to Figure 5.7 summarize the results based on the OHSUMED collection. In this section, I will analyze and discuss findings gathered from the experiments based on the CLEF-IP collection. Please note that the general trends reported here are indicative for both collections used.

As can be seen in diagrams of Figure 5.2 and Figure 5.3, generally, baseline RR resulted in only modest improvements of retrieval performance over baseline PRP, considering the maximum recall reached. In fact, the performance of the two became similar at a page-size of 320 with a recall of 0.63, when the team size was six. However, the diagram in Figure 5.4 indicates that my ILP approach was successful in improving the retrieval performance over both baselines. Considering the maximum recall reached, the improvement was 11%.

However, to obtain a better understanding from the gathered data, in each diagram, I exemplary annotated data points that represent a recall-level of 0.50. This led to the following observations.

Firstly, if one considers a uniform team size of three, we can see that employing the ILP approach led to reduced efforts of the team (compared to the baselines), since less documents needed to be examined to reach the same recall-level. For example, in the diagram of Figure 5.2, we can see that baseline PRP required each team member to examine 200 documents to reach a recall-level of 0.50. Using baseline RR (Figure 5.3), each team member examined 160 documents and finally, using my ILP approach (Figure 5.4), team members examined only 120 documents each to reach recall-level 0.50.

Secondly, if one considers a uniform page-size of 120 documents per team member, the results reveal that by employing the ILP approach, less team members needed to be involved in the search task to reach the same level of recall. For example, in the diagram of Figure 5.2, we see that a team of 5 reached recall-level 0.50 if each member examined 120 documents. Baseline RR (Figure 5.3) allowed a team of 4 reaching this recall-level, and finally, employing the ILP approach (Figure 5.4) reduced the required team size to 3.

To summarize these two observations and to be able to provide a more generic description of the team's efforts, I attempted to implicitly capture the session-costs that resulted from reaching a recall-level of 0.50 (see Table 5.6). I assumed uniform costs required by users to inspect a document. Hence, the sum of documents inspected by the team relates to the team's efforts and, thus, represents an implicit measure (see definition 4.6).

As can be seen in Table 5.6, generally, the PRP approach resulted in highest efforts which most likely resulted from the overlap between the result-pages generated using this baseline. This is improved by both, baseline RR and ILP approach. However, as can been seen in Table 5.6, the ILP approach resulted in the lowest efforts which is also expressed by the highest effectiveness in terms of retrieval performance (see diagrams in Figure 5.2 to Figure 5.7).

**Table 5.6:** Comparison of efforts required to reach a recall level of 0.05.

| Uniform Parameter | Efforts using PRP | Efforts using RR | Efforts using ILP |
|---|---|---|---|
| **Page-size = 120** | (team-size = 5)<br>600 documents | (team-size = 4)<br>480 documents | (team-size = 3)<br>360 documents |
| **Team-size = 3** | (page-size = 200)<br>600 documents | (page-size = 160)<br>480 documents | (page-size = 120)<br>360 documents |

Lastly, I was interested in the gain that resulted from adding further team members to the collaborative search task. This is depicted in Figure 5.8 to Figure 5.13. We see the percentage gain of recall that resulted from increasing the team size. Clearly, adding a second team member led to the largest gain of recall. This applies for all baselines/approaches employed and both test-collections used. Additionally, we can also see that adding further team members to the collaborative search task only creates a marginal increase of gain of recall. This general trend applies for all baselines and also the ILP approach does not change this trend. However, the ILP approach resulted in the highest gains of recall compared to the baselines. This indicates that the ILP approach leverages the effective involvement of team members in the collaborative search task.

### 5.2.4 Significance Tests

To confirm (or disconfirm) the statistical significance of the results, I performed a paired Student's t-test which has been developed by William Sealy Gosset (1876-1937).

Whereas the normal distribution describes a real-valued random variable in a full population, the t-distribution describes this random variable in sampled sub-populations drawn from the full population. Hypothesis tests based on the t-distribution are denoted with t-tests. A single measure of a sampled sub-population, the so called t-statistic, is compared against the t-distribution. In detail, it is checked whether or not the t-statistic is within the region of rejection which is defined by the confidence level ( $\alpha$) and the degrees of freedom (*df*). The t-distribution depends on the number $n$ of samples. For $n \longrightarrow \infty$, the t-distribution becomes the normal distribution.

#### 5.2.4.1 Test Methodology

The paired Student's t-test assesses whether the means of two samples (i.e. experimental results) are statistically different from each other, i.e. are drawn from two different populations. In the case of my experiments, the analysis is based on related measurements, that is, samples are recall values of documents sets created using different result division algorithms. This characteristic justified the use of the paired Student's t-test and thus, only the differences of those paired measurements as well as their standard deviations needed to be considered.

Let *topics* be the set of collection topics and $n = |topics|$, meaning that $n$ also describes the number of samples. Given two sets of paired measurements $X_i$ and $Y_i$, where $i$ identifies a particular collection topic, then $\bar{X}$ and $\bar{Y}$ denote the means (average values) of these sets of measurements. Moreover, let $\widehat{X_\iota} = (X_i - \bar{X})$ and $\widehat{Y_\iota} = (Y_i - \bar{Y})$. A test statistic, with $df = n - 1$ being the degrees of freedom, is computed according to[17]:

$$t_{statistic} = (\bar{X} - \bar{Y}) \sqrt{\frac{n(n-1)}{\sum_i^n \left(\widehat{X_\iota} - \widehat{Y_\iota}\right)^2}} \qquad (7.1)$$

[17] Tutorial on Student's T-Test to be found at: http://projectile.sv.cmu.edu/research/public/talks/t-test.htm

Two sampled sub-populations belong to the same full population, if the value of the corresponding t-statistic (computed using equation 7.1), is less than that of $t_{critical}$, that is, the t-distribution value $t_{distribution}(\alpha, df)$.

If the quantity of a t-statistic exceeds these critical values, $|t_{statistic}| > t_{critical}$, the two sampled sub-population do not belong to the same full populations and thus, the measured performance improvements are not random, but significant improvements.

5.2.4.2 Test Results

For the test of statistical significance, I considered recall-values that resulted from the page-size of fifty because this is the typical page-size employed in patentability searches (see table 2 in [Joho et al. 2010]).

I tested the performance improvements resulting from the baseline RR over the baseline PRP as well as the performance improvements resulting from the ILP approach over the baseline RR. Table 5.7 and Table 5.9 both cover results based on the test-collection CLEF-IP. Statistics were computed based on 143 degrees of freedom and a confidence level of $\alpha$ = 0.01. Table 5.8 and Table 5.10 both cover results based on the test-collection OHSUMED. Statistics were computed based on 67 degrees of freedom and a confidence level of $\alpha$ = 0.01. The statistics presented by these tables include means and standard deviations as well as the values of t-critical and the t-statistic.

As one can see in the tables, the performance improvements are highly significant for most of the settings. In case of usage of the OHSUMED collection, I could not confirm the significance of the results for two cases (see Table 5.9 and Table 5.10). It is interesting to note that the significance of performance improvements of baseline RR over baseline PRP decreases with increasing team size. This can be observed for both test collections used (see Table 5.7 and Table 5.9). Conversely, the significance of performance improvements of ILP approach over baseline RR increased with increasing team size (see Table 5.8 and Table 5.10).

Generally, the high significance of the results from the high number of samples and the resulting degrees of freedom. Moreover, it should be mentioned that statistical significance only indicates that the observed improvements are not random. The statistical significance also doesn’t necessarily mean that the improvements translate into an improved IR application for users. Therefore, the results presented in this chapter only represent potential or best-case improvements of retrieval effectiveness.

**Table 5.7:** Baseline **RR** compared against baseline **PRP** based on using **CLEF-IP**

| Team size | Mean PRP | Std. Dev. PRP | Mean RR | Std. Dev. RR | t-critical | t-statistic |
|---|---|---|---|---|---|---|
| **2** | 0,2923 | 0,1495 | 0,3221 | 0,1671 | 2,61 | **5.27** |
| **3** | 0,3329 | 0,1574 | 0,3820 | 0,1734 | 2,61 | **6,41** |
| **4** | 0,3555 | 0,1671 | 0,4065 | 0,1791 | 2,61 | **6,39** |
| **5** | 0,3790 | 0,1674 | 0,4191 | 0,1768 | 2,61 | **4,97** |
| **6** | 0,3961 | 0,1696 | 0,4226 | 0,1807 | 2,61 | **3,14** |

**Table 5.8:** **ILP** approach compared against baseline **RR** based on using **CLEF-IP**

| Team size | Mean RR | Std. Dev. RR | Mean ILP | Std. Dev. ILP | t-critical | t-statistic |
|---|---|---|---|---|---|---|
| **2** | 0,3221 | 0,1671 | 0,3368 | 0,1693 | 2,61 | **3,91** |
| **3** | 0,3820 | 0,1734 | 0,4084 | 0,1785 | 2,61 | **5,18** |
| **4** | 0,4065 | 0,1791 | 0,4463 | 0,1888 | 2,61 | **7,15** |
| **5** | 0,4191 | 0,1768 | 0,4791 | 0,1857 | 2,61 | **9,78** |
| **6** | 0,4226 | 0,1807 | 0,5026 | 0,1865 | 2,61 | **12,5** |

**Table 5.9:** Baseline **RR** compared against baseline **PRP** based on using **OHSUMED**

| Team size | Mean PRP | Std. Dev. PRP | Mean RR | Std. Dev. RR | t-critical | t-statistic |
|---|---|---|---|---|---|---|
| **2** | 0,4909 | 0,1825 | 0,5511 | 0,1762 | 2,65 | **6,08** |
| **3** | 0,5302 | 0,1771 | 0,5888 | 0,1643 | 2,65 | **5,51** |
| **4** | 0,5460 | 0,1805 | 0,5990 | 0,1645 | 2,65 | **4,32** |
| **5** | 0,5532 | 0,1790 | 0,5961 | 0,1543 | 2,65 | **3,30** |
| **6** | 0,5621 | 0,1799 | 0,5879 | 0,1571 | 2,65 | 1,73 |

**Table 5.10: ILP** approach compared against baseline **RR** based on using **OHSUMED**

| Team size | Mean RR | Std. Dev. RR | Mean ILP | Std. Dev. ILP | t-critical | t-statistic |
|---|---|---|---|---|---|---|
| 2 | 0,5511 | 0,1762 | 0,5662 | 0,1850 | 2,65 | 2,34 |
| 3 | 0,5888 | 0,1643 | 0,6335 | 0,1781 | 2,65 | **5,85** |
| 4 | 0,5990 | 0,1645 | 0,6671 | 0,1747 | 2,65 | **6,75** |
| 5 | 0,5961 | 0,1543 | 0,6869 | 0,1743 | 2,65 | **8,87** |
| 6 | 0,5879 | 0,1571 | 0,7022 | 0,1712 | 2,65 | **9,25** |

### 5.2.5 Limitations

Certainly, several limitations are attached to my results. I analyzed synchronous collaboration because it was in focus in prior CIR research [Soulier et al. 2013] [Chirag Shah et al. 2010] [Jeremy Pickens et al. 2008] [Foley and Smeaton 2009] [Joho et al. 2009]. This choice, however, allowed me to be somewhat comparable with these prior experiments. Thus, I also found trends in the results that have been identified before. For example, as observed in [Joho et al. 2009], we could see that the performance was improved as the team size increased and that the benefit of an extra member was the largest on the second member. Also, my results confirm the observation that a simple Round Robin based result division, called SS5 in [Joho et al. 2009], did not result in substantial performance improvements considering the maximum recall reached. However, I was successful in improving retrieval performance using the novel ILP approach. A further limitation is that the results might not be generalizable to other search tasks types, as I focused on professional, i.e. recall-oriented, search tasks. For example, Web-search tasks, that are known to be precision oriented, my not significantly benefit from the developed result division strategy.

However, despite these limitations, my results allow for confirmation of my initial research hypothesis, and, along with results obtained from prior experiments, support the view that IR can benefit from systems specifically designed to support collaboration in the back-end rather than make users divert systems designed for single-user usage.

# 6 Summary of Contributions and Outlook

In professional practice, IR is often performed in collaboration by teams that utilize a broad set of tools and services that are not specifically designed for collaborative usage. Thus, professionals typically perform their information and collaboration activities in a heterogeneous environment of networked users, tools and information objects.

Ordering documents is a difficult but important task for IR systems optimizing result lists as response to users' queries. This is an even greater difficulty in collaborative search sessions performed in such heterogeneous environments. The objective of the research presented in this dissertation was to develop system-based CIR support for focused and (potentially) geographically distributed teams that aim at resolving a shared information need.

To approach the research objective of this dissertation, I conducted an extensive literature review covering the fields related to this dissertation (see chapter 2). To this end, I collected the most important (informal) definitions of concepts that are used in this dissertation. I continued with reviewing and summarizing empirical field works that captured collaborative IR activities of work teams in real-world settings of different professional domains. Those works identified how users behave and interact in various situations and conditions. Such works were observational in the sense that they identified the various entities involved in IR activities and factors that are likely to have an influence. The findings of these works were, e.g., summarized in conceptual models, such as the one synthesized by me and presented in section 2.3.4.2. However, those works provided only limited guidance to design systems that support collaboration in IR.

To better detail the specific subject of my research, I continued with a systematic analysis of sub-fields in the area of CIR. Recommender systems as well as collaborative re-ranking systems were both classified as systems supporting only implicit collaboration because collaboration here covers the consideration of other users' historical interactions as a source of evidence for document relevance. Social search includes both, explicit collaboration and implicit collaboration but serves the purpose of satisfying the information need of a single person only who is willing or able to involve a larger social network to satisfy that need. The subject of this dissertation was collaborative search, i.e., two or more people who share the same information need and explicitly set out together to satisfy that need [G. Golovchinsky et al. 2008].

I continued with a structured survey on collaborative search support systems. This was structured according to the 3-C Model [Teufel 1995], i.e., the systems' support of communication, coordination, and cooperation between team members. Moreover, the difference between front-end mediated collaboration (awareness information and communication means integrated in the user interface) and back-end mediated collaboration (algorithmic processing of team members' information activities in the search engine) was discussed according to [G. Golovchinsky et al. 2008]. Furthermore, I develop a schema for classifying software tools aiming to support collaborative information activities. The reviewed research was classified in accordance to the developed schema. Finally, I discussed methodologies and problems with IR evaluation and described how these methodologies were adopted towards CIR evaluation.

In chapter 3 of this dissertation, I presented an approach complementing prior research in the area of CIR. Previous research approaches were either observational or experimental. That is, observational works were based on empirical field studies and aimed at capturing the CIR activities at the various stages of the search process. Those works are helpful to describe how people interact and behave under various circumstances. Conversely, in experimental works, study participants were provided with a set of tools implementing various collaboration services and they were asked to perform some predefined tasks. Such works mainly aimed at assessing collaborative tools or setups.

My approach differed from those in terms of methodology. I developed a formal, theoretically sound model for supporting a team during collaborative performance of IR activities in the heterogeneous environments of today's professional practice. Based on a user study, the technical prerequisites of the CIR setting in focus of this dissertation have been captured. The findings of this study led to a conceptual system model that covers the technical environment in which CIR is performed. Moreover, an informal model describing the process of CIR support in such environments has been developed. Based on the conditions defined by the informal models, I introduced a formal cost model and developed a ranking principle, denoted with cPRP, which is a generalization of the well-known PRP and leads to cost-optimal ranking solutions in presence of a team working towards a shared goal.

The classical PRP minimizes costs by ranking according to decreasing values of expected relevance, i.e., the ranking is performed with respect to cost-optimality considering an individual searcher. However, I proposed a ranking strategy for collaborative sessions that considers the trade-off between estimated relevance and estimated information activities of team mates and resulting document redundancy. Whereas an individual user is likely to be satisfied by documents addressing the information need expressed by his or her queries, in the context of team-work, this criterion alone might not be satisfactory. Collaborative searchers share discovered information [Talja 2002] and hence, the cPRP posits documents at higher ranks that are likely to satisfy the shared information need and at the same time have not been discovered yet within the team. This means that in collaborative sessions, the cost-optimal order of documents under the PRP can be different to that of the cPRP and this difference is influenced by the team's information activities.

The cost model developed in this dissertation allowed the derivation of Activity Suggestions. That is, a general criterion that describes optimum collaboration strategies in IR as the solution of an ILP. It is important to note that the developed criterion is *declarative*, i.e., it describes how such an optimum is characterized, but it does not explicitly define the way how this optimum can be computed. This is conceptually distinct to previous *imperative* approaches where it was hypnotized that some concrete scoring function [Jeremy Pickens et al. 2008] or algorithm [Soulier et al. 2013] will result in better retrieval performance for a team.

A prototypical implementation of Activity Suggestions demonstrated the practicability of the developed formal criterion and was described in chapter 4. The prototype was used for a quantitative evaluation in chapter 5. I demonstrated the application of my optimum criterion by means of search result division in two professional search tasks. The influence of a changing team size was studied as well as influence of a changing number of documents examined by each team member. Results yielded improvements of potential retrieval effectiveness of recall-oriented search tasks. That is, generally, my ILP approach allowed a team finding more or as many relevant documents, as the baselines did. I was also able to show that my approach can reduce the efforts required by a team to reach a certain recall-level, so that a team needs to examine less documents or, alternatively, less team members need to be involved in the search task.

It is important to note that the contribution of this dissertation is not the provision of a novel search result division technique even though this was in focus of the experimental part of this dissertation. The utilization of the developed optimum criterion for search result division is one possible use case and was chosen for two reasons: Firstly, to demonstrate feasibility in a comprehensible use case, and, secondly, to be comparable with previous CIR research where search result division was one main topic of investigation [Soulier et al. 2013] [Chirag Shah et al. 2010] [Jeremy Pickens et al. 2008] [Foley and Smeaton 2009] [Joho et al. 2009].

However, the main idea of the developed optimum criterion in section 3.4 was to provide a declarative model for CIR. That is, a model that describes formally how optimum collaboration *"looks like"*. This was achieved by formulating CIR as an ILP which is novel in the field of IR. Depending on the specific CIR application, there might be different strategies for approximating the solution of that ILP. To this end, the chosen baselines in the experiments can be considered as rather rough approximations of the solution of the ILP. Moreover, other algorithms than the chosen baselines are thinkable which could provide more precise approximations of the solution. However, the employment of a numerical solver creates a very precise solution of the same problem and, hence, the corresponding experimental results outperformed the baselines.

## 6.1 Answers to Research Questions

In this section, I briefly summarize the answers to the research questions listed in chapter 1.

- RQ1: *What constitutes a collaborative environment in professional real-world settings used to perform information searching and sharing activities within teams?*

Using an empirical user study, presented in chapter 3, it could be found that today's collaboration environments can be best described as heterogeneous environment where each team member uses personally preferred software applications. It must be assumed that team members use their own personal configuration of software tools for the different information and collaboration activities. Collaborative users perform their information activities independently, i.e., loosely coupled, synchronously or asynchronously.

- RQ2: *How can a team be supported during a collaborative search session? Specifically, how can information search systems be enhanced to reflect team member's information activates?*

Most IR systems that are employed in professional work routines are designed for individual use and, thus, search results are optimized towards an individual. Using a decision theoretic approach, I generalized the PRP to situations where several team members work together. The resulting formal model optimizes abstract costs resulting from information activities considering a whole team. This allowed for describing optimum collaboration strategies in IR as the solution of an ILP which is solvable (and implementable) using numerical methods.

- RQ3: *To which extend does this support increase the potential retrieval effectiveness of the collaborative search tasks?*

Experiments presented in this dissertation showed that the developed model can be integrated in an IR system that is utilized by several team members. Because the research interest of this dissertation addressed professional search tasks, as it is typical for the IP domain, recall-oriented search tasks were setup to study potential retrieval effectiveness. By comparing the developed approach against proven techniques that served as baselines, it could be shown that the developed approach allowed for significant improvements of retrieval effectiveness for a team of searchers.

## 6.2 Future Research Directions

While the contributions summarized above are important ones, future research will need to explore how the achieved improvements of retrieval performance can be translated into real user benefit. The approach presented in this dissertation can be considered as the foundation of an enhancement of today's collaboration environments that consists of tools and interfaces designed for individual usage. Extending this environment with a service dedicated to CIR support would provide the potential for an increase of retrieval effectiveness as indicated by my experiments. Such a service might incorporate the ILP approach presented in this dissertation in different ways, e.g., document sets estimated using the ILP approach could be recommended to users or visually highlighted in results presented to users. However, the design and evaluation of suitable user-interfaces as well as the integration of the developed CIR support into the utilized tools of today’s practice were left to future work.

Moreover, in this dissertation, the feasibility of Activity Suggestions were demonstrated using search result division among team members. In other stages of the Information Dialog, e.g. Access, Activity Suggestions could be used to distribute candidate query terms among team members. However, this task was also left to future work.

The research presented in this dissertation was based on two main ideas. Firstly, the generalization of the PRP and, secondly, employing a decision theoretic approach. The following sections briefly discuss future directions of research that vary these two main ideas.

### 6.2.1 Generalization of other Ranking Principles

Starting point of the model developed in this dissertation was the PRP. It was chosen because it represents the de-facto standard implemented by most search engines, and because its cost-optimality can be proofed formally. However, other ranking principles could also provide such a starting point for the modelling process.

Several approaches addressed the main weakness of the probability ranking principle, that is, the assumed independence between documents. It has often been argued that an examined document potentially changes the user’s information need which, instead of being static, needs to be considered as a “*moving target*” [Fuhr 2008].

For example, one popular strategy is to provide users with documents that are relevant but contain minimal similarity to previous selected items. For example, the MMR approach, ***Maximal Marginal Relevance*** [Carbonell and Goldstein 1998], introduced parameter λ to balance the similarity between candidate document $d$ and query $q_i$ and the similarity between this candidate document $d$ and documents $d'$ ranked at earlier positions. It ranks documents according to decreasing values of

$$\lambda sim(q_i, d) + (1 - \lambda) \max_{d' \in I(q_i)} sim(d, d') \tag{8.1}$$

In this equation, $q_i$ is a query, $I(q_i)$ is the subset of documents previously inspected by a user and thus selected from $R(q_i)$. $sim$ is a similarity metric as it is often used in document retrieval and defined in, e.g., the Vector Space Model [Salton and McGill 1983].

One way of generalizing this to a team situation would be to consider documents already inspected by the whole team. More formally, we would define $d'$ as being an element of $d' \in I(q_i) \cup I(q_{\neg i})$. However, open questions are attached to this approach, such as, how to identify documents $I(q_{\neg i})$ or estimate the knowledge about $I(q_{\neg i})$. Eventually, such an approach would be more appropriate towards search tasks that aim at diversity instead of recall. Only the latter one was the research interest in this dissertation.

### 6.2.2 Non-Linear Cost Models

The approach studied in this dissertation was based on a linear cost model which considered a sum of costs multiplied by the probability of that cost occurring. The authors of [Kahneman and Tversky 1979] presented results of a series of laboratory experiments involving hypothetical choices that showed how humans actually assess gains or losses associated with decisions as well as the likelihood of those events. These findings are, for example [Levy 1994]:

1) People tend to think in terms of gains and losses rather than in terms of their net assets, i.e., they take choices by considering deviations from a reference point.
2) People treat gains differently than losses, i.e., individuals tend to be risk-averse with respect to gains and risk-acceptant with respect to losses.
3) Gains are also treated differently than losses in that losses dominate larger than gains.

Kahneman and Tversky proposed to describe the expected utility of a decision using two weighing functions. First, to adjust the conditional gains (or losses) associated with a decision, and second, to reflect the subjective probabilities. By employing the same notation as used in equation 3.1, Kahneman and Tversky summarized the expected utility that people aim to maximize as follows.

$$EU(d) = \sum_{k} u\big(\mathfrak{L}(\alpha|\omega_k)\big)\, v(P(\omega_k|d)) \tag{8.2}$$

In this equation, $u$ is a function that assigns a value to costs incurred by a decision[18]. The function $v$ is a probability weighting function and captures the idea that people tend to overestimate small probability events, but underestimate large probabilities. This yields a non-linear cost-function.

Future research would have to answer questions such as how these weighting functions could be quantified in the context of IR and how such an adjustment of the cost model could improve the team support mechanism presented in this dissertation. However, such an approach has the potential advantage of considering the actual decision behavior of real users.

[18] Note that here, $u$ also converts losses into gains

# References


John Adcock and Jeremy Pickens. 2008. FXPAL collaborative exploratory video search system. In *Proceedings of the 2008 international conference on Content-based image and video retrieval*. ACM, 551–552.

Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on* 17, 6 (2005), 734–749.

T.M. Amabile et al. 2001. Academic-practitioner collaboration in management research: A case of cross-profession collaboration. *Academy of Management Journal* (2001), 418–431.

Avi Arampatzis and Jaap Kamps. 2008. A study of query length. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 811–812.

Simon Attfield, Ann Blandford, and Stephann Makri. 2010. Social and interactional practices for disseminating current awareness information in an organisational setting. *Information processing & management* 46, 6 (2010), 632–645.

Leif Azzopardi. 2009. Query side evaluation: an empirical analysis of effectiveness and effort. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 556–563.

Leif Azzopardi, Kalervo Järvelin, Jaap Kamps, and Mark D. Smucker. 2011. Report on the SIGIR 2010 workshop on the simulation of interaction. In *ACM SIGIR Forum*. ACM, 35–47.

Leif Azzopardi, Maarten de Rijke, and Krisztian Balog. 2007. Building Simulated Queries for Known-Item Topics. In *Proceedings of the 30st annual international ACM SIGIR conference on Research and development in information retrieval*. 455–462.

Leif Azzopardi, Wim Vanderbauwhede, and Hideo Joho. 2010. Search system requirements of patent analysts. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 775–776.

Ricardo Baeza-Yates and José A. Pino. 1997. A first step to formally evaluate collaborative work. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work: the integration challenge*. ACM, 56–60.

Marcia J. Bates. 2011. *Understanding information retrieval systems: management, types, and standards*, CRC Press.

M.J. Bates. 1979a. Idea tactics. *Journal of the American Society for Information Science* 30, 5 (1979), 280–289.

M.J. Bates. 1979b. Information search tactics. *Journal of the American Society for information Science* 30, 4 (1979), 205–214.

Thomas Beckers, Sebastian Dungs, Norbert Fuhr, Matthias Jordan, Sascha Kriewel, and Vu Tuan Tran. 2012. ezdl: An interactive search and evaluation system. In *SIGIR 2012 Workshop on Open Source Information Retrieval*. 9–16.

N.J. Belkin, C. Cool, A. Stein, and U. Thiel. 1995. Cases, scripts, and information-seeking strategies: on the design of interactive information retrieval systems. *Expert systems with applications* 9, 3 (1995), 379–395.

N.J. Belkin, R.N. Oddy, and H.M. Brooks. 1982a. ASK for information retrieval: Part I. Background and theory. *Journal of documentation* 38, 2 (1982), 61–71.

N.J. Belkin, R.N. Oddy, and H.M. Brooks. 1982b. ASK for information retrieval: Part II. Results of a design study. *Journal of documentation* 38, 3 (1982), 145–164.

Paul N. Bennett et al. 2012. Modeling the impact of short-and long-term behavior on search personalization. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 185–194.

B. Billerbeck, F. Scholer, H.E. Williams, and J. Zobel. 2003. Query expansion using associated queries. In *Proceedings of the twelfth international conference on Information and knowledge management*. 2–9.

Frank Blackler. 1995. Knowledge, knowledge work and organizations: an overview and interpretation. *Organization studies* 16, 6 (1995), 1021–1046.

Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. 2014a. Collaborative Information Seeking in Professional Work-Settings: A Study of Equipment Utilization. *Datenbank-Spektrum* 14, 1 (2014), 29–38.

Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. 2014b. ezDL: Collaborative Information Seeking and Retrieval in a Heterogeneous Environment. *Computer* 47, 3 (2014), 32–37.

Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. 2013. Supporting Collaborative Information Seeking and Searching in Distributed Environments. In *Proc. LWA 2013 Conf.*

Dario Bonino, Alberto Ciaramella, and Fulvio Corno. 2010. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information* 32, 1 (2010), 30–38.

Abraham Bookstein. 1983. Outline of a general probabilistic retrieval model. *Journal of Documentation* 39, 2 (1983), 63–72.

Chahrazed Bouhini, Mathias Géry, and Christine Largeron. 2014. Integrating user's profile in the query model for Social Information Retrieval. In *Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on*. IEEE, 1–2.

J.M. Bowers and S.D. Benford. 1990. *Studies in computer supported cooperative work: theory, practice and design*, North-Holland Publishing Co.

H. Bruce, R. Fidel, A.M. Pejtersen, S. Dumais, J. Grudin, and S. Poltrock. 2003. A comparison of the collaborative information retrieval behaviour of two design teams. *The New Review of Information Behaviour Research* 4, 1 (2003), 139–153.

H. Bukvova. 2010. Studying Research Collaboration: A Literature Review. (2010).

Katriina Byström and Preben Hansen. 2002. Work tasks as units for analysis in information seeking and retrieval studies. In *Proceedings of the 4th International Conference on Conceptions of Library & Information Science (CoLIS4)*. 239–252.

K. Byström and P. Hansen. 2005. Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science and Technology* 56, 10 (2005), 1050–1061.

Giacomo Cabri, Letizia Leonardi, and Franco Zambonelli. 1999. Supporting cooperative WWW browsing: a proxy-based approach. In *16th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP 2008)*. IEEE Computer Society, 138–138.

Huanhuan Cao et al. 2009. Context-aware query classification. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 3–10.

R. Capra, J. Arguello, A. Chen, K. Hawthorne, G. Marchionini, and L. Shaw. 2012. The ResultsSpace collaborative search environment. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*. 435–436.

Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 335–336.

Ed H. Chi. 2008. Information seeking can be social. *Information Seeking Support Systems* (2008).

M. Cole et al. 2009. Usefulness as the criterion for evaluation of interactive information retrieval. *Proc. HCIR* (2009), 1–4.

Kate Crawford, Helen M. Hasan, Leoni Warne, and Henry Linger. 2009. From traditional knowledge management in hierarchical organizations to a network centric paradigm for a changing world. (2009).

Anita Crescenzi and Robert Capra. 2013. Collaboration in search processes and results sharing. *CSCW 2013* (2013).

Jonathon N. Cummings. 2004. Work groups, structural diversity, and knowledge sharing in a global organization. *Management science* 50, 3 (2004), 352–364.

Thomas H. Davenport and Laurence Prusak. 1998. *Working knowledge: How organizations manage what they know*, Harvard Business Press.

P.J. Denning and P. Yaholkovsky. 2008. Getting to we. *Communications of the ACM* 51, 4 (2008), 19–24.

B. Dervin and M. Nilan. 1986. Information needs and uses. *Annual review of information science and technology* 21 (1986), 3–33.

P. Dourish and V. Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. 107–114.

C.A. Ellis, S.J. Gibbs, and G. Rein. 1991. Groupware: some issues and experiences. *Communications of the ACM* 34, 1 (1991), 39–58.

D. Ellis. 1989. A behavioural approach to information retrieval system design. *Journal of documentation* 45, 3 (1989), 171–212.

Felix Engel. 2015. *Expertensuche in semantisch integrierten Datenbeständen*. Fernuniversität in Hagen.

Martin J. Eppler and Jeanne Mengis. 2004. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The information society* 20, 5 (2004), 325–344.

B.M. Evans and E.H. Chi. 2008. Towards a model of understanding social search. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 485–494.

R. Fidel, H. Bruce, A. Pejtersen, S. Dumais, J. Grudin, and S. Poltrock. 2000. Collaborative information retrieval. *The New Review of Information Behaviour Research* 1, 1 (2000), 235–247.

L. Fitzpatrick and M. Dent. 1997. Automatic feedback using past queries: social searching? In *ACM SIGIR Forum*. 306–313.

Colum Foley and Alan F. Smeaton. 2009. Synchronous collaborative information retrieval: Techniques and evaluation. In *Advances in Information Retrieval*. Springer, 42–53.

C. Foley and A.F. Smeaton. 2010. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *Information processing & management* 46, 6 (2010), 762–772.

J. Foster. 2006. Collaborative information seeking and retrieval. *Annual review of information science and technology* 40, 1 (2006), 329–356.

L. Fuchs, U. Pankoke-Babatz, and W. Prinz. 1995. Supporting cooperative awareness with local event mechanisms: The GroupDesk system. In *Proceedings of the 1995 European Conference on Computer-Supported Cooperative Work (ECSCW'95)*. 247–263.

Norbert Fuhr. 1999. A decision-theoretic approach to database selection in networked IR. *ACM Transactions on Information Systems (TOIS)* 17, 3 (1999), 229–249.

Norbert Fuhr. 2008. A probability ranking principle for interactive information retrieval. *Information Retrieval* 11, 3 (2008), 251–265.

Norbert Fuhr. 1992. Probabilistic models in information retrieval. *The Computer Journal* 35, 3 (1992), 243–255.

Susan Gauch, Mirco Speretta, Aravind Chandramouli, and Alessandro Micarelli. 2007. User profiles for personalized information access. In *The adaptive web*. Springer, 54–89.

Gene Golovchinsky, John Adcock, Jeremy Pickens, Pernilla Qvarfordt, and Maribeth Back. 2008. Cerchiamo: a collaborative exploratory search tool. *Proceedings of Computer Supported Cooperative Work (CSCW)* (2008).

G. Golovchinsky, J. Pickens, and M. Back. 2009. A taxonomy of collaboration in online information seeking. *Arxiv preprint arXiv:0908.0704* (2009).

G. Golovchinsky, P. Qvarfordt, and J. Pickens. 2008. Collaborative information seeking. *Information Seeking Support Systems* (2008).

Michael D. Gordon and Peter Lenk. 1991. A utility theoretic examination of the probability ranking principle in information retrieval. *JASIS* 42, 10 (1991), 703–714.

Michael D. Gordon and Peter Lenk. 1992. When is the probability ranking principle suboptimal? *JASIS* 43, 1 (1992), 1–14.

Jonathan Grudin. 1994. Computer-supported cooperative work: History and focus. *Computer* 27, 5 (1994), 19–26.

Quanquan Gu and Jiawei Han. 2011. Towards feature selection in network. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 1175–1184.

Jörg M. Haake and Brian Wilson. 1992. Supporting collaborative writing of hyperdocuments in SEPIA. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 138–146.

Uri Hanani, Bracha Shapira, and Peretz Shoval. 2001. Information filtering: Overview of issues, research and systems. *User Modeling and User-Adapted Interaction* 11, 3 (2001), 203–259.

P. Hansen. 2011. *Task-based Information Seeking and Retrieval in the Patent Domain: Processes and Relationships*. Tampere University Press.

P. Hansen and K. Järvelin. 2005. Collaborative information retrieval in an information-intensive domain. *Information Processing & Management* 41, 5 (2005), 1101–1119.

Matthias Hemmje, Adelheit Stein, and Heinz-Dieter Boecker. 1996. A multidimensional categorization of information activities for differential design and evaluation of information systems. (1996).

William Hersh et al. 2000. Do batch and user evaluations give the same results? In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 17–24.

William Hersh, Chris Buckley, TJ Leone, and David Hickam. 1994. OHSUMED: An interactive retrieval evaluation and new large test collection for research. In *SIGIR 94*. Springer, 192–201.

Morten Hertzum and Annelise Mark Pejtersen. 2000. The information-seeking practices of engineers: searching for documents as well as for people. *Information Processing & Management* 36, 5 (2000), 761–778.

M. Hertzum and A.M. Pejtersen. 2000. The information-seeking practices of engineers: searching for documents as well as for people. *Information Processing & Management* 36, 5 (2000), 761–778.

David Hunt, Long Nguyen, and Matthew Rodgers. 2012. *Patent searching: Tools & techniques*, John Wiley & Sons.

A. Hust. 2005. Query expansion methods for collaborative information retrieval. *Informatik-Forschung und Entwicklung* 19, 4 (2005), 224–238.

J. Hyldegard. 2006. *Between individual and group-exploring group members' information behavior in context*, Department of Information Studies, Royal School of Library and Information Science.

J. Hyldegard. 2009. Uncertainty Dimensions of Information Behaviour in a Group Based Problem Solving Context. *Nordic Journal of Information Literacy in Higher Education* 1, 1 (2009).

Bernard J. Jansen, Amanda Spink, and Tefko Saracevic. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management* 36, 2 (2000), 207–227.

Kalervo Järvelin. 2009. Interactive relevance feedback with graded relevance and sentence extraction: simulated user experiments. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2053–2056.

K. Järvelin and T.D. Wilson. 2003. On conceptual models for information seeking and retrieval research. *Information Research* 9, 1 (2003), 9–1.

Robert Johansen. 1988. *Groupware: Computer support for business teams*, The Free Press.

Hideo Joho, Leif A. Azzopardi, and Wim Vanderbauwhede. 2010. A survey of patent users: an analysis of tasks, behavior, search functionality and system requirements. In *Proceedings of the third symposium on Information interaction in context*. ACM, 13–24.

Hideo Joho, David Hannah, and Joemon M. Jose. 2008. Comparing collaborative and independent search in a recall-oriented task. In *Proceedings of the second international symposium on Information interaction in context*. ACM, 89–96.

Hideo Joho, David Hannah, and Joemon M. Jose. 2009. Revisiting IR techniques for collaborative search strategies. In *Advances in Information Retrieval*. Springer, 66–77.

Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society* (1979), 263–291.

Arvind Karunakaran, Patricia Ruma Spence, and Madhu C. Reddy. 2010. Towards a model of collaborative information behavior. In *2nd International Workshop on Collaborative Information Seeking-ACM Conference on Computer Supported Cooperative Work. ACM*.

Ryan Kelly and Stephen J. Payne. 2014. Collaborative web search in context: a study of tool use in everyday tasks. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 807–819.

Heikki Keskustalo, Kalervo Järvelin, Ari Pirkola, Tarun Sharma, and Marianne Lykke. 2009. Test collection-based ir evaluation needs extension toward sessions–a case of extremely short queries. In *Information Retrieval Technology*. Springer, 63–74.

Youngho Kim, Jangwon Seo, and W. Bruce Croft. 2011. Automatic boolean query suggestion for professional search. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 825–834.

C.P. Klas, S. Kriewel, and M. Hemmje. 2008. An experimental system for adaptive services in information retrieval. In *Proceedings of the 2nd International Workshop on Adaptive Information Retrieval (AIR 2008)*.

Donald Ervin Knuth. 1998. *The art of computer programming: sorting and searching*, Pearson Education.

Sascha Kriewel and Norbert Fuhr. 2010. Evaluation of an adaptive search suggestion system. In *Advances in Information Retrieval*. Springer, 544–555.

Rainer Kuhlen. 1991. Information and pragmatic value-adding: Language games and information science. *Computers and the Humanities* 25, 2-3 (1991), 93–101.

Rainer Kuhlen. 2004. Informationsethik–Formen des Umgangs mit Wissen und Information in elektronischen Räumen. *Erscheint* 9 (2004), 2004.

C.C. Kuhlthau. 1991. Inside the search process: Information seeking from the user's perspective. *Journal of the american society for information science* 42, 5 (1991), 361–371.

Paul Landwich. 2012. *Unterstützung von Information-Retrieval-Prozessen durch Informationsdialogkontext-Verwaltung*. Hagen, Fernuniv., Diss., 2012.

P. Landwich, C.P. Klas, and M. Hemmje. 2009. Catching the user-logging the information retrieval dialogue. In *SIGIR 2009 Workshop: Understanding the User, Boston, USA*.

P. Landwich, T. Vogel, C.P. Klas, and M. Hemmje. 2009. Supporting patent retrieval in the context of innovation-processes by means of information dialogue. *World Patent Information* 31, 4 (2009), 315–322.

Victor Lavrenko. 2010. Introduction to Probabilistic Models for Information Retrieval. In *proceedings of the 33rd International ACM SIGIR conference on Research and Development in Information Retrieval edited by Hsin-His Chen, Efthismis N. Efthimiadis, Jacques Savoy, Fabio Crestani Lugano and Stephane Marehand-Maillet, Association for Computing Machinery, New York*. 905.

Jack Levy. 1994. An introduction to prospect theory. *Avoiding Losses/Taking Risks: Prospect Theory and International Conflict, edited by Barbara Farnham* (1994), 7–22.

G. Linden, B. Smith, and J. York. 2003. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE* 7, 1 (2003), 76–80.

Scott London. 1995. Collaboration and community. *Retrieved July* 12 (1995), 2011.

Byron Long and Ronald Baecker. 1997. A Taxonomy of Internet Communication Tools. In *WebNet*.

G. Marchionini. 1997. *Information seeking in electronic environments*, Cambridge Univ Pr.

Adam Mathes. 2004. Folksonomies-cooperative classification and communication through shared metadata. *Computer Mediated Communication* 47, 10 (2004), 1–13.

Joseph E. McGrath, Holly Arrow, Deborah H. Gruenfeld, Andrea B. Hollingshead, and Kathleen M. O'Connor. 1993. Groups, tasks, and technology The effects of experience and change. *Small Group Research* 24, 3 (1993), 406–420.

Göran Melin and Olle Persson. 1996. Studying research collaboration using co-authorships. *Scientometrics* 36, 3 (1996), 363–377.

Angelos Mitrelis, Giannis Tsakonas, and Christos Papatheodorou. 2008. Using qualitative research methods for the improvement of collaborative information searching tools design. (2008).

Stefano Mizzaro. 1997. Relevance: The whole history. *JASIS* 48, 9 (1997), 810–832.

Meredith Ringel Morris. 2013. Collaborative search revisited. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1181–1192.

M.R. Morris. 2008. A survey of collaborative web search practices. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. 1657–1660.

M.R. Morris. 2007. Interfaces for collaborative exploratory web search: Motivations and directions for multi-user design. In *Proceedings of ACM SIGCHI 2007 Workshop on Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction*. 9–12.

M.R. Morris and E. Horvitz. 2007. SearchTogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. 3–12.

M.R. Morris, J. Lombardo, and D. Wigdor. 2010. WeSearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 401–410.

M.R. Morris, J. Teevan, and S. Bush. 2008. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. 481–484.

H. Naderi, B. Rumpler, and J.M. Pinon. 2007. An efficient collaborative information retrieval system by incorporating the user profile. *Adaptive Multimedia Retrieval: User, Context, and Feedback* (2007), 247–257.

Jakob Nielsen. 1994. Usability inspection methods. In *Conference companion on Human factors in computing systems*. ACM, 413–414.

Ikujirō Nonaka and Hirotaka Takeuchi. 1995. *The knowledge-creating company: How Japanese companies create the dynamics of innovation*, Oxford university press.

D. Warner North. 1968. A tutorial introduction to decision theory. *Systems Science and Cybernetics, IEEE Transactions on* 4, 3 (1968), 200–210.

Jay F. Nunamaker Jr and Minder Chen. 1990. Systems development in information systems research. In *System Sciences, 1990., Proceedings of the Twenty-Third Annual Hawaii International Conference on*. IEEE, 631–640.

Vicki L. O'Day and Robin Jeffries. 1993. Information artisans: patterns of result sharing by information searchers. In *Proceedings of the conference on Organizational computing systems*. ACM, 98–107.

Dorab Patel and Scott D. Kalter. 1993. Low overhead, loosely coupled communication channels in collaboration. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW'93*. Springer, 203–218.

Jeremy Pickens, Gene Golovchinsky, Chirag Shah, Pernilla Qvarfordt, and Maribeth Back. 2008. Algorithmic mediation for collaborative exploratory search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 315–322.

J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. 2008. Algorithmic mediation for collaborative exploratory search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 315–322.

Vassilis Plachouras, Ben He, and Iadh Ounis. 2004. University of Glasgow at TREC 2004: Experiments in Web, Robust, and Terabyte Tracks with Terrier. In *TREC*.

S. Poltrock, J. Grudin, S. Dumais, R. Fidel, H. Bruce, and A.M. Pejtersen. 2003. Information seeking and sharing in design teams. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*. 239–247.

Steven Poltrock, Jonathan Grudin, Susan Dumais, Raya Fidel, Harry Bruce, and Annelise Mark Pejtersen. 2003. Information seeking and sharing in design teams. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*. ACM, 239–247.

Paul Prekop. 2002. A qualitative study of collaborative information seeking. *Journal of Documentation* 58, 5 (2002), 533–547.

M.C. Reddy and B.J. Jansen. 2008. A model for understanding collaborative information behavior in context: A study of two healthcare teams. *Information Processing & Management* 44, 1 (2008), 256–273.

M.C. Reddy and P.R. Spence. 2008. Collaborative information seeking: A field study of a multidisciplinary patient care team. *Information processing & management* 44, 1 (2008), 242–255.

Stephen E. Robertson. 1977. The probability ranking principle in IR. *Journal of documentation* 33, 4 (1977), 294–304.

Stephen E. Robertson and David A. Hull. 2000. The TREC-9 Filtering Track Final Report. In *TREC*. 25–40.

Stephen Robertson and Hugo Zaragoza. 2009. *The probabilistic relevance framework: BM25 and beyond*, Now Publishers Inc.

Stephen Robertson, Hugo Zaragoza, and Michael Taylor. 2004. Simple BM25 extension to multiple weighted fields. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. ACM, 42–49.

Giovanna Roda, John Tait, Florina Piroi, and Veronika Zenz. 2010. CLEF-IP 2009: retrieval experiments in the Intellectual Property domain. In *Multilingual Information Access Evaluation I. Text Retrieval Experiments*. Springer, 385–409.

T. Rodden. 1991. A survey of CSCW systems. *interacting with Computers* 3, 3 (1991), 319–353.

Jesus Alberto Rodriguez Perez, Stewart Whiting, and Joemon M. Jose. 2011. CoFox: A visual collaborative browser. In *Proceedings of the 3rd international workshop on Collaborative information retrieval*. ACM, 29–32.

Markus Rohde and David Williamson Shaffer. 2003. Us, ourselves, and we: Thoughts about social (self-) categorization. *ACM SIGGROUP Bulletin* 24, 3 (2003), 19–24.

Nicholas C. Romano Jr, J.F. Nunamaker Jr, D. Roussinov, and Hsinshun Chen. 1999. Collaborative information retrieval environment: Integration of information retrieval with group support systems. In *System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on*. 10–pp.

Jennifer E. Rowley. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science* (2007).

Ian Ruthven. 2003. Re-examining the potential effectiveness of interactive query expansion. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 213–220.

Gerard Salton and Michael J. McGill. 1983. Introduction to modern information retrieval. (1983).

Mark Sanderson. 2010. *Test collection based evaluation of information retrieval systems*, Now Publishers Inc.

Mark Sanderson and Justin Zobel. 2005. Information retrieval system evaluation: effort, sensitivity, and reliability. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 162–169.

T. Saracevic. 1997. The stratified model of information retrieval interaction: Extension and applications. In *PROCEEDINGS OF THE ANNUAL MEETING-AMERICAN SOCIETY FOR INFORMATION SCIENCE*. 313–327.

Tefko Saracevic. 1975. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science* 26, 6 (1975), 321–343.

Tefko Saracevic. 2007. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology* 58, 13 (2007), 2126–2144.

Tefko Saracevic. 1996. Relevance reconsidered. In *Proceedings of the second conference on conceptions of library and information science (CoLIS 2)*. ACM Press, 201–218.

Abigail J. Sellen, Rachel Murphy, and Kate L. Shaw. 2002. How knowledge workers use the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 227–234.

C. Shah. 2011. A framework for supporting user-centric collaborative information seeking. In *ACM SIGIR Forum*. 88–88.

C. Shah and R. González-Ibáñez. 2011. Evaluating the synergic effect of collaboration in information seeking. In *SIGIR'11: Proceedings of the 34th annual international ACM SIGIR conference on Research and development in information retrieval, July 24*. 24–28.

C. Shah and R. González-Ibáñez. 2010. Exploring information seeking processes in collaborative search tasks. *Proceedings of the American Society for Information Science and Technology* 47, 1 (2010), 1–7.

Chirag Shah. 2014. Evaluating collaborative information seeking–synthesis, suggestions, and structure. *Journal of Information Science* (2014), 0165551514530651.

Chirag Shah and Roberto González-Ibáñez. 2012. Spatial context in collaborative information seeking. *Journal of Information Science* 38, 4 (2012), 333–349.

Chirag Shah, Jeremy Pickens, and Gene Golovchinsky. 2010. Role-based results redistribution for collaborative information retrieval. *Information processing & management* 46, 6 (2010), 773–781.

C. Shah and G. Marchionini. 2010. Awareness in collaborative information seeking. *Journal of the American Society for Information Science and Technology* 61, 10 (2010), 1970–1986.

C. Shah, J. Pickens, and G. Golovchinsky. 2010. Role-based results redistribution for collaborative information retrieval. *Information processing & management* 46, 6 (2010), 773–781.

Ben Shneiderman. 1992. *Designing the user interface: strategies for effective human-computer interaction*, Addison-Wesley Reading, MA.

A.F. Smeaton, H. Lee, C. Foley, and S. McGivney. 2007. Collaborative video searching on a tabletop. *Multimedia systems* 12, 4 (2007), 375–391.

Barry Smyth, Jill Freyne, Maurice Coyle, Peter Briggs, and Evelyn Balfe. 2004. I-SPY—Anonymous, Community-Based Personalization by Collaborative Meta-Search. In *Research and Development in Intelligent Systems XX*. Springer, 367–380.

B. Smyth, E. Balfe, J. Freyne, P. Briggs, M. Coyle, and O. Boydell. 2004. Exploiting query repetition and regularity in an adaptive community-based web search engine. *User Modeling and User-Adapted Interaction* 14, 5 (2004), 383–423.

D.H. Sonnenwald, K.L. Maglaughlin, and M.C. Whitton. 2004. Designing to support situation awareness across distances: an example from a scientific collaboratory. *Information Processing & Management* 40, 6 (2004), 989–1011.

Laure Soulier, Chirag Shah, and Lynda Tamine. 2014. User-driven system-mediated collaborative information retrieval. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 485–494.

Laure Soulier, Lynda Tamine, and Wahiba Bahsoun. 2013. A collaborative document ranking model for a multi-faceted search. In *Information Retrieval Technology*. Springer, 109–120.

P.R. Spence, M.C. Reddy, and R. Hall. 2005. A survey of collaborative information seeking practices of academic researchers. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*. 85–88.

A. Spink, T. Wilson, D. Ellis, and N. Ford. 1998. Modeling users' successive searches in digital environments. *D-lib Magazine* 4, 4 (1998).

JohnI. Tait. 2014. An Introduction to Professional Search. In Georgios Paltoglou, Fernando Loizides, & Preben Hansen, eds. *Professional Search in the Modern World*. Lecture Notes in Computer Science. Springer International Publishing, 1–5.

S. Talja. 2002. Information sharing in academic communities: types and levels of collaboration in information seeking and use. *New Review of Information Behavior Research* 3, 1 (2002), 143–159.

S. Talja and P. Hansen. 2006. Information sharing. *New directions in human information behavior* (2006), 113–134.

E. Taylor-Powell, B. Rossing, and J. Geran. 1998. Evaluating collaboratives: reaching the potential. *Madison, WI: University of Wisconsin—Cooperative Extension* (1998).

Stephanie Teufel. 1995. *Computerunterstützung für die Gruppenarbeit*, Addison-Wesley.

M.B. Twidale, D.M. Nichols, and C.D. Paice. 1997. Browsing is a collaborative process. *Information Processing & Management* 33, 6 (1997), 761–783.

Michael Twidale and David Nichols. 1998. A survey of applications of CSCW for digital libraries: Technical Report CSEG/4/1998. (1998).

Robert Villa, Nicholas Gildea, and Joemon M. Jose. 2008. A study of awareness in multimedia search. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 221–230.

Ellen M. Voorhees. 2000. Variations in relevance judgments and the measurement of retrieval effectiveness. *Information processing & management* 36, 5 (2000), 697–716.

J. Walkerdine and T. Rodden. 2001. Sharing searches: developing open support for collaborative searching. In *Human-computer Interaction: INTERACT'01: IFIP TC. 13 International Conference on Human-Comupter Interaction, 9th-13th July 2001, Tokyo, Japan*. 140.

Jun Wang and Jianhan Zhu. 2009. Portfolio theory of information retrieval. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 115–122.

Ryen W. White, Paul N. Bennett, and Susan T. Dumais. 2010. Predicting short-term interests using activity-based search context. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 1009–1018.

Ryen W. White, Susan T. Dumais, and Jaime Teevan. 2009. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. ACM, 132–141.

Ryen W. White, Ian Ruthven, Joemon M. Jose, and C.J. Van Rijsbergen. 2005. Evaluating implicit feedback models using searcher simulations. *ACM Transactions on Information Systems (TOIS)* 23, 3 (2005), 325–361.

James Wilsdon and others. 2011. *Knowledge, networks and nations: global scientific collaboration in the 21st century*, The Royal Society.

T.D. Wilson. 1999. Models in information behaviour research. *Journal of documentation* 55, 3 (1999), 249–270.

Zhen Yue, Shuguang Han, and Daqing He. 2014. Modeling search processes using hidden states in collaborative exploratory web search. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 820–830.

Guido Zuccon and Leif Azzopardi. 2010. Using the quantum probability ranking principle to rank interdependent documents. In *Advances in information retrieval*. Springer, 357–369.

Guido Zuccon, Leif Azzopardi, and CJ van Rijsbergen. 2011. The interactive PRP for diversifying document rankings. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 1227–1228.

Guido Zuccon, Leif Azzopardi, and CJ van Rijsbergen. 2010. Has portfolio theory got any principles? In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 755–756.

# Lebenslauf

| | | |
|---|---|---|
| **Berufliche Tätigkeiten** | 07/2003 – jetzt SAP SE<br>• Softwareentwickler im Bereich PI Technology Core Plattform | Walldorf |
| | 12/2001 – 06/2003 Konrad-Zuse-Zentrum für Informationstechnik<br>• Wissenschaftlicher Mitarbeiter im Bereich Scientific Visualization | Berlin |
| **Studium** | 10/2007 – 12/2011 FernUniversität in Hagen<br>• Studium der Elektro- und Informationstechnik | Hagen |
| | 10/1997 – 11/2001 Fachhochschule für Technik und Wirtschaft<br>• Studium der Technischen Informatik | Berlin |
| **Schulbildung** | 08/1983 – 06/1996<br>• Allgemeine Hochschulreife (Abitur) | Berlin |

## Eidesstattliche Erklärung

Hiermit versichere ich, dass die vorliegende Dissertation *„System-Mediated Support of Explicit Collaboration in Information Retrieval“* selbständig und ohne unerlaubte fremde Hilfe angefertigt und keine anderen als die in der Dissertation angegebenen Hilfsmittel benutzt habe.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Die vorliegende Dissertation hat zuvor keiner anderen Stelle zur Prüfung vorgelegen.

Thilo Böhm