Examining the Effectiveness of Collaborative Search Engines

Moraes Gomes, F.

DOI
10.4233/uuid:9f8f9ffc-c3d9-4583-8903-25889233a95b

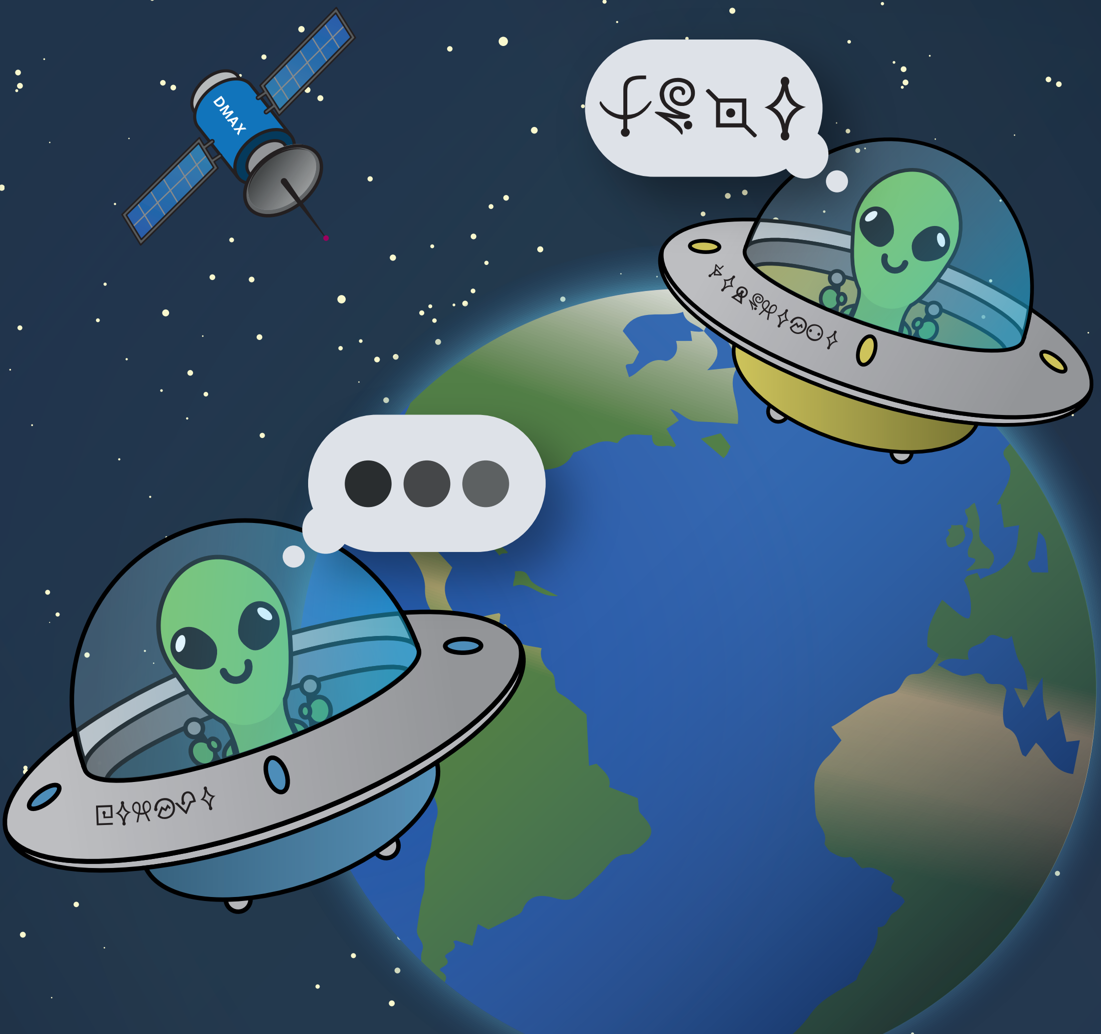
Publication date
2022

Document Version
Final published version

Citation (APA)
Moraes Gomes, F. (2022). *Examining the Effectiveness of Collaborative Search Engines*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:9f8f9ffc-c3d9-4583-8903-25889233a95b

Examining the Effectiveness of Collaborative Search Engines

Felipe Moraes Gomes



Examining the Effectiveness of Collaborative Search Engines

Examining the Effectiveness of Collaborative Search Engines

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen
op donderdag 24 februari 2022 om 15.00 uur

door

Felipe MORAES GOMES

Master of Science in Computer Science,
Universidade Federal de Minas Gerais, Brazilië,
geboren te Divinópolis, Minas Gerais, Brazilië.

Dit proefschrift is goedgekeurd door de

promotor: Prof. dr. ir. G.J.P.M Houben
promotor: dr. C. Hauff

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. G.J.P.M. Houben,	Technische Universiteit Delft
Dr. C. Hauff,	Technische Universiteit Delft
Onafhankelijke leden:	
Prof. dr. P.S. Cesar Garcia,	Technische Universiteit Delft
Prof. dr. ir. D. Hiemstra	Radbound Universiteit
Prof. dr. U. Kruschwitz	University of Regensburg, Germany
Dr. L. Azzopardi,	University of Strathclyde, Scotland
Dr. V. Murdock,	Amazon, United States of America
Prof. dr. M.M. Specht,	Technische Universiteit Delft, reservelid

SIKS Dissertation Series No. 2022-10
The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems. This thesis has been supported by NWO project LACrOSSE (612.001.605).

Published and distributed by: Felipe Moraes Gomes
E-mail: moraesfgomes@gmail.com

Keywords: Collaborative Search, Search Engines, Sensemaking

Printed by: Proefschriftenprinten.nl

Cover: Dr. David Maxwell
Typeset in *Livvic* under the terms of the *Open Font License*

Style: TU Delft House Style, with modifications by Moritz Beller
`https://github.com/Inventitech/phd-thesis-template`

ISBN: 978-94-6384-307-2

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

Aos meus pais, Divina e João.

On a given day, a given circumstance, you think you have a limit. And you then go for this limit and you touch this limit, and you think, ‘Okay, this is the limit’. And so you touch this limit, something happens and you suddenly can go a little bit further.

Ayrton Senna

Contents

Acknowledgments

This PhD thesis is the result of the effort and support of several people to whom I am extremely grateful.

First of all, the first words to express my special appreciation and gratitude are to my parents, Divina and João. This PhD thesis is dedicated to you. Also, I would like to thank my very kind sisters, Kellen and Keroleine, and my nephew and niece, Pedro e Valentina. Thank you for allowing me to pursue this path, for all your unconditional love, and all for the support that you gave me until today. We have been through a lot in these last four years. I love you!

Primeiramente, as minhas primeiras palavras para expressar meus agradecimentos vão para meus pais, Divina e João. Essa tese de doutoado é dedicada à vocês. Também, gostaria de agradecer as minhas irmãs, Kellen e Keroline, e os meus sobrinhos, Pedro e Valentina. Obrigado por permitir que eu seguisse esse caminho, pelo seu amor incondional e por todo suporte que vocês me deram até hoje. Passamos por muita coisa nesses últimos quatro anos. Eu te amo!

A PhD student's work can be seen as a solitary path, but it is only a fruitful path if you have a caring supervisory team as I had. I want to express my deepest gratitude to my supervisory team Geert-Jan Houben, and Claudia Hauff, for giving me the opportunity to work under their supervision. I am very thankful for your continuous support and encouragement to continue to improve my work. Claudia, I am grateful for your scientific guidance, constructive criticism and wholehearted support when I needed it. Thank you for challenging me and making my thoughts into meaningful outcomes. Geert-Jan, thank you for the insightful conversations with critical and constructive comments. You were constantly enforcing me to think about the bigger picture.

I would like to thank the members of my defence committee: Prof. Pablo Cesar Garcia, Prof. Djoerd Hiemstra, Prof. Udo Kruschwitz, Prof. Marcus Specht, Dr. Leif Azzopardi, and Dr. Vanessa Murdock, for accepting to be part of my committee and providing me with valuable feedback on this thesis.

There are many people to that I need to pay a special tribute when looking back at my journey. Starting from the current and former members of the Web information systems group at Delft University of Technology: Jie Yang, Ioannis Petros Samiotis, Sihang Qiu, Sepideh Mesbah, Yue Zhao, Shahin Sharifi, Asterios Katsifodimos, Nava Tintarev, Andrea Mauri, Achilleas Psyllidis, Dimitrios Bountouridis, Shabnam Najafian, Dan Davis, Christos Koutras, Carlo van der Valk, Pavel Kucherbaev, Alessandro Bozzon, Christoph Lofi, Ujwal Gadiraju, Agathe Balayn, Sara Salimzadeh, Georgios Siachamis, Peide Zhu, Manuel Valle Torre, Guanliang Chen, Andra Lonescu, Mónica Marrero, Tamara Brusik, and Daphne Stephan. Also, my collaborators at WIS Delft and externals: Sindunuraga R. Putra, Killian Grashoff, Daan Rennings, and Sheikh M. Sarwar. A special thank you for supporting me in the difficult times, especially at the end of this journey, for hearing all my complaints: Arthur Camara, Gustavo Penha, David Maxwell, Tim Draws, and Nirmal Roy.

This journey would not have been possible without the love and support of friends. My gratitude is not enough, but special thanks go to Thales, Luis Eduardo, Priscylla, Thaís, Gabriela, Ana Luisa, Sara, Laís, Jeanderson, Jordan, and Nicholas. Thales, thank you for being receptive during all the times I went to Dublin, and for motivating me and demotivating me at the same time during our almost weekly calls. Dudu, thank you for being so kind, and even with days of delayed replies to my Whatsapp messages, you always made time to hear me out. Pri, thank you for being this honest person, and caring for me; the distance will not stop us to think about each other. Nicholas, thank you for the best caipirinhas in Amsterdam. To my kind housemates that I had during the PhD, Breno and Javad, thank you for your patience and the long chats.

Last, but certainly not least, I own my deepest gratitude to Renan. I am so fortunate to have had you by my side, supporting me in this challenging path in the last two years. Thank you for cheering me up in my vulnerable moments, for holding my hands when I needed strength, and for being patient every day with my boring complaints about this thesis.

To everyone I mentioned here (and maybe I forgot some names), I hope I have not neglected to show you how important you are to me in my life. I am so blessed to have all of you in my life. Thank you for everything.

Amsterdam, The Netherlands

1

Introduction

1.1 Motivation

Searching for information is considered an essential skill in our daily lives. With roots in the library sciences, modern Information Retrieval (IR) systems, such as Google and Bing (hereon referred to as search engines), are today computerised, providing nearly instantaneous access to huge volumes of information [16].

Search engines are used by individuals that have a particular information need in mind. For example, given a circumstance, a user will realise the existence of an information need that could arise from a potential knowledge gap, a lack of understanding of how to connect their current knowledge, or a conflict of evidence that they need to resolve. Known as the Anomalous State of Knowledge (ASK) [19], a user will, in turn, take this information need and attempt to formulate it as a query (e.g., a textual representation of what they are looking for) before issuing the query to a search engine. Following this, the search engine assesses the documents within its index and returns a ranked list of documents given the user's query. The user then examines the ranked list and hopefully will encounter the answers they are looking for that satisfy their information need.

The information need that a user develops can often be represented as a search task– a goal that the user needs to achieve whilst searching [72]. For example, a search task where a user searches for Thai restaurants in Amsterdam would have one possible goal of finding the best affordable Thai restaurant to dine at. On the other hand, a more complex search task may require a student to use academic search engines such as Google Scholar to search for and identify several academic articles that argue for or against the given subject (i.e., the subject that they are examining). Thus, search tasks can range from the simplistic (i.e., identifying known facts such as 'How old is King Willem-Alexander?') to the more complex (i.e., spanning multiple stages with the search engine, such as a literature search) [5].

Although completing those search tasks is often seen as a solitary activity, searching in collaboration with others is deemed useful or necessary in many situations [100]. Often these situations involve a group of people working together to complete a task where search occurs in part of the task or during the entire task. Examples include travel planning, online shopping, looking for health related information, planning birthday parties, working on a group project, or finding a house to buy [58, 76, 93, 116, 170]. Additionally, collaboration can also occur in a work context when people are using search as part of their daily job [78]. The main benefit of collaborative search (CSE) is that a group of people can execute a search task more effectively (e.g., number of hotels found) and more efficiently (e.g., time taken to plan a trip) than when a search task is executed individually [124]. In particular, a good CSE experience seeks for obtaining a *"true synergy among diverse participants in creating solutions or strategies through the synergistic interactions of a group of people"* [71]. In the work context, CSE may have implications such as cost of searching given that the collaborator may not choose to work with one or more colleagues and may have different roles in the company/organisation. As part of one of our user studies, we surveyed 305 crowdworkers about their prior collaborative search experiences. In Table 1.1, we selected a few of those collaborative search experiences categorised by collaborative task types on the left column.

Task type	**Examples of collaborative search experiences.**
Travel planning ✈	*- Planning and booking a trip to London with overnight stay and travel.* *- Looking for hotels and flights to Portugal.* *- Me and my wife planning a trip to Turkey booking hotels flights and transport to airport.*
Shopping 🛒	*- Family members looking for a certain product.* *- Husband and wife researching old book prices.* *- Friend wanted a dash cam with specific features, and I helped locate a number of products to compare.*
Health information ✚	*- Searching for diabetes related treatments.* *- Looking for info/reviews of surgical doctors.*
Social planning 🎂	*- My wife and I were looking at things to do in the Lake District. So I was searching while the wife was telling me what she wants to look for and do.* *- Myself and daughter searching for party ideas for my son.* *- Looking for movies that are currently out.*
Learning 📕	*- Group of students researching human biology.* *- A group of students working on a writing assignment and sharing search results/findings.* *- Working on a group project in high school with other students.*
Real estate 🏠	*- My sister and I were searching for a flat.* *- A new house with my spouse.* *- We spent some time searching together our next property before moving house.*

Table 1.1: Previous collaborative experiences described by participants in our user study in May 2018 (see Chapter 3 for details of the study).

Despite the popularity of the aforementioned commercial search engines—and open-source search engine frameworks such as Elasticsearch[1], Terrier [108], or Indri [139]—these are not suitable for supporting collaborations amongst a group of users. When searching in collaboration, many users often use these search engines together with other external tools (e.g., e-mails, collaborative editors such as Google Docs or Overleaf, video chat such as Zoom or Microsoft Teams), so that they can be aware of each other's activities, share the information they found, and divide the labour of the task [100]. Having to use more than one tool may often be undesirable as it adds the extra overhead to coordinate the collaboration using more than one tool [102]. Thus, focused on empowering research in CSE, several dedicated CSE systems have been developed over the last 15 years with the objective to support collaborations within an all-in-one system [30, 54, 68, 84, 102, 109, 128, 163, 167]. These collaborations have been categorised according to four dimensions [52]:

Intent: Collaborators' intent can be *implicit*, without necessarily being aware of each other's existence, or *explicit*, where collaborators are aware of each other's activities. An example of implicit collaboration is a recommender system, where past interactions of other users are used to provide recommendations to users with similar profiles. In contrast, in explicit (intentional) collaboration, users explicitly share an information need. In all examples described in Table 1.1, collaborators have an explicit intent.

Concurrency: CSE can occur when collaborators search at the same (*synchronous*) or at different times (*asynchronous*). For instance, synchronous CSE occurs when collaborators aim to complete a task, where the task needs negotiation and communication in order to achieve each user's requirements (e.g., travel planning). Different from asynchronous CSE, the final result of the search task is build up iteratively. For instance, a group of students doing a literature review for their group project, where one group member can improve the results found by other group members on earlier iterations in later iterations. In the examples described in Table 1.1, collaborations can take place synchronously or asynchronously.

Location: Collaborators can be either physically *co-located* or *remote*. In remote collaboration, a CSE system often provides communication channels such as text chat or video chat. For example, remote collaboration can occur in search tasks when collaborators cannot meet personally. In co-located collaboration, users can use their own device such as laptop or smartphones, for instance, in social planning activities, such as watching films to watch as exemplified in Table 1.1.

Mediation: Collaborations can be mediated at either *user-interface (UI)* level or *algorithmic mediation* level. For example, in UI based mediation, the underlying CSE system plays a secondary role, with collaborators primarily making use of user interface elements such as a chat or visual aids to be aware of each others' activity and communicate. In contrast, algorithmic mediation takes place on the CSE system level, where the system contains algorithms that provide strategies to foster collaboration. In the examples described in

[1] https://www.elastic.co/elasticsearch/

Table 1.1, collaborators can be mediated via UI level or algorithmic mediation level. However, a mediation level such as algorithmic mediation may be more suitable for tasks that require users to divide the task before the start of the collaborative search session (e.g., looking for hotels in different regions of Amsterdam).

Over the years, researchers have found that complex search tasks can be executed more effectively (e.g., the number of relevant documents found during the search session), be more efficient (e.g., in terms of time), achieve higher material coverage, and enable higher knowledge gains in an explicit collaborative setting than if conducted in isolation [126, 130]. However, some CSE tasks have not shown to benefit from users collaborating in terms of the amount of relevant information found (e.g., precision) [131]. In fact, one of the challenges in CSE research is to understand the drawbacks of collaboration. In particular, different than single-user search, in synchronous remote CSE, communication and coordination can be an overhead during the search task. Moreover, not enough awareness of each other's activities can hinder completing a search task. **Hence, this thesis focuses on examining when explicit, synchronous, remote CSE is effective.** We focus on synchronous remote CSE due to the increased demand in recent years for remote and synchronous collaborations as people are more distributed across the globe and need to accomplish CSE tasks [7].

To evaluate the effectiveness of CSE systems, a common evaluation methodology in CSE research is designing user experience research (UER) studies (e.g., user studies) [72], which consist of evaluation methodologies from the fields of interactive information retrieval (IIR) and human computer interaction (HCI) [12]. In particular, typically, we measure effectiveness of a CSE system based on performance-based measures related to the outcome of the search part of a task, such as the number of relevant documents found by a group of users during the task [14, 15].

Even though researchers have carefully designed several CSE user studies, there is still conflicting evidence or a lack of evidence on the effectiveness of CSE systems [69, 111, 129–131, 135, 144]. Thus, in this thesis, we focus on examining the effectiveness of CSE systems in two parts. In the first part, we shed light on the effectiveness of CSE to support two group configurations, in turn, namely *group sizes* and *users' roles*.

Group Sizes: Previous collaborative search studies have had a strong focus on groups of two or three collaborators, thus naturally limiting the number of experimental conditions that could increase quickly [3, 23, 24, 55, 63, 64, 69, 102, 103, 111, 126, 129, 135, 136, 144]. Therefore, there is a lack of evidence on the extent to which a CSE system can support group sizes beyond these commonly investigated group sizes. Thus, we study CSE system effectiveness with group size as the primary dependent variable, varying groups sizes of two to six collaborators, with six as our upper bound due to our available resources.

Users' Roles: Roles can determine how a group splits up the search task and determine each group member's function (e.g., one group member is responsible for finding documents and reading and evaluating them, another one for in-depth reading and evaluating them). In particular, when the CSE system assigns a role to each group member, researchers have hypothesised that a group may reduce time spent communicating and coordinating the task and make the search process more efficient and successful than groups without role assignment [111]. However, past user studies have provided contradicting evidence as to the utility of assigned roles in CSE [111, 129, 144]. Thus, we provide more evidence that aims to settle the question of the effectiveness of CSE systems when used by groups with pre-assigned roles versus groups without pre-assigned roles.

In the second part of this thesis, we fix our group configurations, particularly, group sizes are set to up to three people, and group members receive the same role. Then, we turn to a different perspective and focus on examining the effectiveness in two contexts: search as learning and collaborative online shopping. We chose these two contexts because they include everyday collaborative tasks, exemplified in Table 1.1, and because search is a significant part of these contexts.

Search as Learning: In the last decades, for many learners, web search engines are the first step to access and ingest information [22, 106, 146] in their learning activities. Search activities for human learning involve multiple iterations that require cognitive processing and interpretation, often requiring the searcher to spend time scanning/viewing, comparing, and evaluating information. However, web search engines are not built to support users in the search tasks often required in learning situations [54, 59, 88]. When people use search as a learning activity, it can be an individual activity or a collaborative activity (e.g., group projects). Therefore, in this thesis, we tackle the challenge of identifying the impact of web search engines on the (single-search or collaborative search) users ability to learn compared to learning acquired via high-quality learning materials as a baseline.

Collaborative Online Shopping: In collaborative online shopping, a group of people come together to make a decision to purchase a product that meets the various group members' requirements and opinions [3, 170]. While shopping together, search is an important part of the task in order to search for products in a catalogue that are available in an e-commerce website. One important aspect of collaborative shopping is supporting awareness and sharing of knowledge as it can enable a sense of co-presence, which helps groups make a decision that satisfies each group member's requirements and wishes. As search is a significant part of a collaborative online shopping experience, CSE systems are suitable for executing such tasks. However, there is insufficient evidence of how well can CSE systems support a group of users to search for online products together and make a group decision. Hence, we explore the effects of increased awareness and sharing of knowledge (co-presence) using a CSE system in collaborative shopping on the group decision making process.

1.2 Research Questions

This thesis consists of two parts aimed at examining (i) the impact of differences in *group configuration* on CSE system effectiveness (Chapters 3 and 4), and (ii) the effectiveness of CSE systems in support of collaborative tasks in the context of search as learning and collaborative online shopping (Chapters 5 and 6). First, Chapter 2 presents a background overview and our CSE system to enable the research conducted in the rest of this thesis. Finally, in Chapter 7, we conclude the thesis and discuss directions for future work.

Here, we describe the main research questions for each chapter. Then, in each chapter, we outline more fine-grained subquestions.

1.2.1 Effectiveness on Supporting Group Configurations

One group configuration that has been often considered as a constant in CSE research is the *size* of the collaborating group. Two studies that conducted experiments increasing *group sizes* found evidence that larger groups can reach higher search effectiveness in a recall-oriented task [70, 130], though with diminishing returns (each increase in group size decreases the rate the search effectiveness increases). However, one of these studies was conducted with simulated users [70], overlooking the potential increase in cognitive load of real users. Cognitive load refers to the impact capacity of the user to hold on to information along with the session. Another study that aims to understand the impact of group sizes took place in a co-located setting, which may not suffer from the same coordination overhead as remote studies [130]. Therefore, there are still two open questions (i) to what extent the simulation findings carry over to user studies, and (ii) to what extent the currently existing CSE support group sizes beyond the commonly studied sizes of two or three.

In this part of the thesis, the first research question that we answer is following:

RQ1 What is the impact of group size on CSE in terms of retrieval effectiveness and search behaviours?

To answer **RQ1**, in Chapter 3, we present a crowd-sourcing study with 305 participants based on prior best principles with group size as the main dependent variable, investigating group sizes of two, four and six collaborating searchers conducted across three recall-oriented search topics. We find that most prior simulation-based results on the impact of group size on behaviour and search effectiveness do not hold in our user study with several hundred crowd-workers. More importantly, we do not observe diminishing returns (measured in recall) when increasing group sizes from two to six collaborators.

Another group configuration that has been of interested in previous work are the roles (e.g., functions) of each group member. Intuitively, assigning roles to each group member seems to reduce the amount of communication necessary in a group, thus leaving searchers with more time to focus on the task at hand [111]. However, despite the intuitive appeal of role-based CSE, prior user studies have revealed conflicting evidence [111, 129, 144]. In these studies, the experimental design did not consider the comprehensive comparison of role-based CSE with individual searchers and artificially created groups from individual searchers. Comparing with artificial groups is vital in order to observe the existence of the benefits of CSE in terms of the *synergic effect* (i.e., the set of documents found by groups cover more relevant documents than documents found by merging the set of documents found by individual users). With a full experimental design (i.e., considering all possible experimentation conditions), we can answer our second research question:

RQ2 What are the benefits of role-based CSE in recall-oriented tasks in terms of search retrieval effectiveness, search behaviours, and communication patterns?

To this end, in Chapter 4, we conduct another user study (120 participants) with three recall-oriented tasks across four conditions: (i) individual searchers; (ii) artificial groups created by aggregating the search logs of pairs of individual searchers; (iii) pairs of collaborating searchers without assigned roles; and (iv) pairs of collaborating searchers with assigned roles. We do not find evidence of the benefits of collaborating groups with assigned roles regarding search effectiveness. In particular, artificial groups (i.e., pairs of individual searchers) who do not benefit from communication and coordination perform at the same level of recall as collaborative groups. In addition, we find that groups with assigned roles exhibit search behaviours that indicate their compliance with the given role distribution: group members tasked with exploring the search space spend less time reading documents and issue more queries. Differently, group members tasked with exploring the set of documents pre-selected by other group members do indeed spend more time reading documents. Overall, despite the apparent differences in search behaviours, we do not find evidence that groups with pre-assigned roles are more beneficial than groups without roles.

1.2.2 Supporting Search as Learning and Online shopping Contexts

Researchers have explored the support of CSE in many contexts where collaboration seems beneficial. In Table 1.1, we show a few collaborative experiences where collaboration is used in daily lives. In the second part of this thesis, we explore the effectiveness of CSE systems supporting collaborative tasks in the contexts of search as learning and collaborative shopping. We chose these two contexts as the effectiveness of CSE systems does not depend on search effectiveness alone. Instead, in a search as learning context, effectiveness is measured by learning gains, which is the product of a learning task, and in collaborative online shopping, the focus is on group decision making process and the impact on the process to complete a collaborative shopping using a CSE system. So, first, we look at search as learning context, and we aim to answer the following research question:

RQ3 Is (individual and collaborative) search a viable alternative to instructor-designed learning?

To answer **RQ3**, in Chapter 5, we present a user study with 151 participants and measure learning gains through a *vocabulary learning* task. In order to build a competitive baseline for possible learning gains, we use as a baseline how much learning a user can acquire from instructor-designed learning materials (online lecture videos). In particular, our experimental conditions consist of one individual search condition, two individual search conditions combined with instructor-designed learning materials, and one collaborative search condition. We find our participants in the instructor-designed learning condition (watching online lecture videos) to have higher learning gains than participants in the search condition. Additionally, we find that the *combination* of instructor-designed learning and searching to learn leads to significantly higher learning gains than the instructor-designed learning condition without the combination with search. Finally, we find that CSE as a learning activity does not result in increased learning gains.

Next, we look at the context of collaborative online shopping. In this context, as search plays a significant role during the task, CSE systems can be used to support a group of people shopping together. In particular, a CSE system can support a group during a shopping task that requires the group to make a decision given each member's constraints and opinions. In this scenario, the goal is to provide a better CSE experience that enables and increases the sense of co-presence. Co-presence is the feeling of awareness of each other's activity which is often perceived in collaborative physical shopping. By enabling co-presence in a CSE system, researchers have found that it can decrease the communication overhead among members and improve group coordination during the decision making process through navigation mechanisms [33, 49, 156, 166, 170]. A baseline navigation mechanism is separate navigation, where users' navigation is independent of each other (e.g., none of the users' actions affects the others' activities). Another navigation mechanism is collaborative navigation, which allows a group to navigate and share information via an intermediate interface (e.g., the user search results page (SERP) updates every time another user issues a new query or paginates the SERP) Thus, in Chapter 6, we answer the following research question:

RQ4 How and when do navigational mechanisms impact the group decision making process in collaborative shopping using a CSE system?

To answer **RQ4**, we present our last user study with 64 participants across two navigation mechanisms, namely, separate and collaborative navigation. We investigate to what extent a group of collaborating users differs in their search and decision making behaviour when faced with either separate or collaborative navigation. We find that groups in the collaborative navigation are more effective in their search space exploration, which means they explore more products in less time. Also, shared navigation causes less usage of typical collaborative search features. Finally, although collaborative navigation increases collaborators' co-presence during the task, we do not find a significant impact of collaborative navigation on purchase and post-purchase perceived satisfaction.

1.3 Main Contributions

In this section, we summarise the main contributions of this thesis.

Conceptual contributions

1. We conduct an overview of the literature of CSE systems focused on the best practices of developing CSE systems and open research directions (Chapter 2).
2. We conduct a comparison of related studies and empirical evaluations of CSE in terms of group size, number of groups, number of search tasks per group, and study type (Chapter 3).
3. We conduct a comparison of related studies of role-based CSE in terms of roles, study type, number of groups, number of search tasks, and division of labour types (Chapter 4).

Empirical contributions

4. We conduct a large-scale experiment with more than 300 crowd-workers on large group sizes and the impact of group size dynamics on users' behaviour and search metrics (Chapter 3).
   (a) We compare the impact of groups size 2, 4, and 6 on CSE system performance.

(b) We compare the impact of CSE system variants based on levels of sharing of knowledge.

5. We conduct a crowd-sourced user study (120 participants) on user's roles to shed light on the effectiveness of pre-assigned user's role in CSE (Chapter 4).

   (a) We compare four experimental conditions: (i) individual searchers; (ii) artificial groups; (iii) pairs without assigned roles; and (iv) pairs with assigned roles.

   (b) We analyse the impact of roles on search performance, search behaviour, searchers' communication patterns across the four conditions.

6. We conduct a crowd-sourced user study with 151 participants and measure learning gains across five experimental conditions (Chapter 5).

   (a) We compare (individual and collaborative) search and instructor-designed learning effects on learning gains.

   (b) We analyse the search behaviour and learning gains in search as learning task.

7. We conduct a virtual lab study[2] (64 participants) to investigate the impact of collaborative navigation on the group decision making process in collaborative shopping tasks (Chapter 6).

   (a) We compare separative and collaborative navigation on stages of the group decision making process.

   (b) We analyse the search behaviour in comparison with the group decision making stages.

Resources

8. We provide an open-source CSE system, `SearchX` (Chapter 2). Our system is modular, easy to extend with plenty of features for experimentation using modern web technologies. The code is available at the following address `https://github.com/searchx-framework/`.

   (a) We support algorithmic mediation components and features that enable efficient use of `SearchX` for crowdsourcing studies.

   (b) We provide a product search user interface with search facets, product filters, and separate and shared navigation mechanisms.

[2] In a virtual lab study, participants are typically invited to take part in the experiment via university social media and internal communication tools. In a crowd-sourced user study, participants are invited via a crowdsourcing platform (e.g., FigureEight).

1.4 Origins

We now list the publications on which the research chapters were based.

Chapter 2 is based on three conference papers:

- Sindunuraga Rikarno Putra, Felipe Moraes, and Claudia Hauff. Searchx: Empowering collaborative search research. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1265–1268. Association for Computing Machinery, 2018.
- Sindunuraga Rikarno Putra, Kilian Grashoff, Felipe Moraes, and Claudia Hauff. On the development of a collaborative search system. In *Design of Experimental Search & Information REtrieval Systems*, pages 76–82, 2018.
- Felipe Moraes and Claudia Hauff. node-indri: Moving the indri toolkit to the modern web stack. In *European Conference on Information Retrieval*, pages 241–245. Springer, 2019.

Chapter 3 is based on the journal paper:

- Felipe Moraes, Kilian Grashoff, and Claudia Hauff. On the impact of group size on collaborative search effectiveness. *Information Retrieval Journal*, 22(5):476–498, 2019.

Chapter 5 is based on the conference paper:

- Felipe Moraes, Sindunuraga Rikarno Putra, and Claudia Hauff. Contrasting search as a learning activity with instructor-designed learning. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 167–176, 2018.

Chapter 6 is based on the workshop paper:

- Felipe Moraes, David Maxwell, and Claudia Hauff. Exploring collaborative navigation support in collaborative product search. In *Proceedings of the SIGIR Workshop on eCommerce*, 2021.

The thesis also benefited from insights gained from the following publication:

- Felipe Moraes, Jie Yang, Rongting Zhang, and Vanessa Murdock. The role of attributes in product quality comparisons. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 253–262, 2020

2

SearchX Framework: Empowering Collaborative Search Research

With the motivation set out, we present in this chapter one of the major contributions of this thesis. We describe the design space of collaborative search engines, and outline the challenges in implementing SearchX*, our implementation of an open-source collaborative search engine—complying with modern web standards.* SearchX *implements essential features of collaborative search as found in the literature. We focused on providing support for modern research needs, such as running crowdsourcing experiments and fast prototyping. We use* SearchX *in the user studies in the remaining chapters of this thesis. We open-sourced* SearchX *at* https://github.com/searchx-framework/searchx*. The contributions of this chapter have been mainly published in [94, 114, 115].*

2.1 Introduction

Web search is generally seen as a solitary activity, as most mainstream technologies are designed for single-user search sessions. However, for a sufficiently complex task, collaboration during the information seeking process is beneficial [48]. A survey by Morris in 2012 [101] has shown that collaborating during search is a common activity, albeit using ad hoc solutions such as email and instant messaging. Morris also found a significant increase in the number of people who collaborate during search at a regular basis, from 0.9% in 2006 to 11% in 2012. This increasing use of CSE has also been reflected in the research community, where CSE has been an active area of research for many years. Workshops that explicitly focus on CSE—and more generally information seeking—have started to appear in 2008 [112] and continue to do so to this day, for instance, [12].

In contrast to single-user search where a number of up-to-date and open-source tools are readily available (e.g., `Terrier` [86], `Anserini` [164], and `Elasticsearch`), the CSE research community had up to 2018 only a single, actively maintained, and open-source option (`Coagmento`, see Table 2.1)—despite the fact that researchers had designed, implemented, and reported a number of other systems over the years [30, 54, 68, 84, 102, 109, 128, 163, 167]. While `Coagmento` provides an extensive collaboration feature set, it requires users to either install a browser plugin or an Android/iOS app, making it less viable for large-scale CSE experiments which are often conducted with crowd workers. Furthermore, we believe as researchers we should have a choice of tooling, instead of relying on a single one.

For these reasons, we have designed and implemented `SearchX`, a CSE system built on modern Web standards, allowing it to be accessed from multiple platforms without the need for user-side installations. We designed `SearchX` specifically for CSE research and provide a comprehensive documentation to enable others to implement and run their own CSE experiments.

Based on these observations, we build a system for collaborative web search that can be used as a starting point for future research. As CSE is still an active research area, we design the system specifically for conducting user studies.

Thus, the main contributions of this chapter are:

- We analyse the findings of existing systems to find current best practices as well as common research directions.

- We describe the design process of implementing `SearchX`, which implements essential features of CSE and accommodates extensions for the research of this thesis and for future research.

Additionally, we have successfully deployed `SearchX` in four CSE user studies using a crowd-sourcing platform and virtual lab user studies. Besides, we have identified at least four user studies that used or extended `SearchX` for a crowd-sourcing experiment [28, 119, 121, 123]. With `SearchX`, we contribute the second open-source system for collaborative Web search, which comes with a comprehensive documentation focusing on modifications to the system.

2.2 Background

CSE is a subset of the more generic field of collaborative information seeking (CIS). Golovchinsky et al. [52] have characterised the collaboration aspect of online CIS along four dimensions: *intent* (explicit or implicit), *mediation* (user interface (UI) or algorithm), *concurrency* (synchronous or asynchronous), and *location* (remote or co-located). Morris [101] suggested two additional dimensions: *role* (symmetric or asymmetric) and *medium* (Desktop or and large touch displays). In our interpretation, CSE is scoped around the collaboration of explicit intent, with the mediation and role dimension explored in designing the system, and the other three dimensions (concurrency, location, medium) describing the potential application scenarios of the system.

Systems for CSE. Our analysis of previously proposed systems in Table 2.1 is limited to those similar to SearchX—systems that support at least *synchronous* and *remote* collaborations. Additionally, we limit the scope to text retrieval systems, since it is the most common use case in web search. One of the first attempts at such system was the design of SearchTogether [102], which focused on supporting awareness, division of labour, and persistence. Paul and Morris [109] built CoSense, an extension for SearchTogether to improve sense-making by providing additional views. Shah et al. [128] built upon the weaknesses of SearchTogether and created Coagmento, which has been analysed for its experimental suitability in [73].

More recent systems were created to explore specific aspects of online collaborations. Golovchinsky et al. [54] designed Querium to better support the collaboration in an exploratory search process, specifically through implementing a shared document history that ranks documents based on relevance feedback. Jesus et al. [68] proposed CoFox to allow remote collaborators share their current window while the local collaborator had an independent window to navigate their searches. Capra et al. [30] designed ResultsSpace to study mostly asynchronous collaborations (though synchronous collaborations are possible too), therefore features for direct communication such as a chat were not added. Yue et al. [167] investigated the search behaviour of users and designed CollabSearch with basic collaborative features for analysing difference in search tactics for single-search users vs CSE users. Leelanupab et al. [84] explored the effectiveness of visual snippets for sense-making by introducing the SnapBoard feature into the CoZpace system. Lastly, Xu et al. [163] developed PairSearch to support collaboration between pairs of users searching the web synchronously with an intention to acquire knowledge about a given topic.

As stated before, most of the listed systems are only described in publications, and not open-sourced (or even available as binaries). As mediation is a vital factor for CSE, we first analysed how mediation was designed in prior works, and used that as a starting point in implementing SearchX. As CSE solutions should require low additional effort compared to single-user search systems [54, 73, 101], we strove to implement features that look familiar to users (who all use web search engines) today.

Designing Mediation. There are two main directions in developing mediation for CSE: *interface mediation* adapts the search interface towards a multi-user context, usually in the form of a shared workspace; *system mediation* directly mediates the collaboration process, mostly through re-ranking of documents [30, 54] or modifying the distribution of documents [102]. Both types of mediation are complementary to each other. The support for collaboration features can be categorised along three lines [47, 127]: *division of* *labour*, *sharing of knowledge*, and *awareness*. Table 2.1 provides a feature comparison of prior works [30, 54, 68, 84, 102, 109, 128, 163, 167] in relation to these concepts. We now elaborate on each one and outline what is implemented in SearchX.

Division of Labour refers to the distribution of work load across collaborators. This division can be left to the user (user-driven) or mediated by the system. The latter can be implemented at the user level through assignment of roles (e.g., prospector and miner roles [111]) or at the document level through assigning different document subsets to domain experts [136]. Prior systems mostly support user-driven division of labour through the provision of communication features. Group chat and document sharing (the explicit recommendation of a document to a collaborator) are two features which have been shown to be favoured by users [102, 128]. Following this, SearchX implements group chat; we argue that document sharing in the sense above can be achieved through the chat feature as well and thus does not warrant a separate UI element.

Sharing of Knowledge refers to the ability to share ideas and information effectively between collaborators [165]. This can be facilitated either through shared workspaces [113], or through the re-ranking of search results based on relevance feedback [46]. Prior systems support sharing of knowledge primarily through providing a shared workspace with features for collectively capturing information. Bookmarking documents (i.e. document saving) and document rating are both relevance feedback mechanisms, with prior systems either implementing one or the other. Document bookmarking promotes shortlisting, which involves forming and refining a shared list of potential resources [73]; document rating provides a finer granularity of feedback, which is needed for algorithmic mediation [30, 54]. Document annotation supports the previous two features by communicating the rationale behind an action [73, 102]. The choice of features largely depends on the experimental setup, therefore SearchX implements all three in a way that toggling individual features is easy. In SearchX document saving is instantiated as *bookmarking* as users are familiar with this concept. In the meantime, Roy et al. [121] has extended SearchX to enable users to highlight pieces of text from search results and save them in a notepad which they may collaboratively edit with their group members.

Awareness is defined as *"the ability to maintain some knowledge about the situation and activities of others"* [85], encompassing knowledge of the workspace and collaborators' actions, as well as the ability to instantaneously notice changes on the work conducted. Prior systems focus on providing lightweight information regarding collaborators' search activities (e.g., query history, and document history colour coded to reflect the user that issued a query or viewed a document) and the overall sense-making process (document metadata and summary of group actions such as previously issued queries and saved snippets). All features were found to be useful in past experiments, except for the document history which Kelly and Payne [73] reported to provide too much information. As group summaries are mostly beneficial for asynchronous sessions [127], SearchX implements query history, document metadata, and colour coding as awareness features.

Table 2.1: Feature comparison of existing remote CSE systems and SearchX (ordered by publication year of the first paper describing the system). A dash indicates that this information is not available. Language and platform abbreviations: JS=JavaScript, BP=Browser Plugin, IE=Internet Explorer, FF=Firefox, GC=Google Chrome. Note that we only list programming languages listed in the respective papers (if no open-source code is available). The Coagmento iOS app is only available in Apple's US app store.

	SearchTogether [102]	CoSense [109]	Coagmento [128]	Querium [54]	CoFox [68]	ResultsSpace [30]	CollabSearch [167]	CoZpace [84]	PairSearch [163]	SearchX
Division of Labour										
Group Chat	✓	✓	✓		✓	✓		✓	✓	✓
Document Sharing	✓		✓	✓					✓	
Sharing of Knowledge										
Bookmarking / Document Saving			✓		✓		✓			✓
Document Rating	✓	✓		✓		✓		✓		✓
Document Annotation	✓	✓	✓					✓		✓
Awareness										
Query History	✓	✓	✓	✓		✓	✓	✓		✓
Document History		✓	✓	✓	✓					
Document Metadata	✓	✓	✓	✓		✓		✓		✓
Group Summary		✓		✓			✓	✓		
Colour Coding		✓					✓			✓
System Mediation	Split Search	–	–	Ranked Doc. History	Shared Window	Re-ranked Search Results	–	–	Screen Sharing	–
Tool Availability										
Functioning	✗	–	✓	–	–	–	–	–	–	✓
Open Source	–	–	✓	–	–	–	–	–	–	✓
Last Update	2009	–	2018	–	–	–	–	–	–	2021
Language	–	–	PHP & JS	JS	–	PHP	–	JS	PHP & JS	JS
Platform	BP (IE)	–	iOS†, Android BP (FF, GC)	Web	Web 6	Web	Web	Web	Web	Web

2.3 System Design

We engineered SearchX as a system that can be easily adopted by other CSE researchers. SearchX adopts common collaborative features found in prior literature works and presents them in a way that is familiar to generic users. Additionally, we emphasise support for modern research needs, specifically support for fast prototyping of new features or interfaces, and support for online studies.

SearchX is designed for the following experimental workflow: a researcher first implements an experimental setup of their user study using SearchX (either relying on existing features, or adding their own). Each study participant accesses the SearchX instance through a designated URL; the browsers Safari, Google Chrome, and Mozilla Firefox in desktop version with JavaScript enabled are supported. For single-user experiments, the system allocates a `sessionId` to each participant, and for the collaborative experiments, the system then allocates a search `sessionId` and `groupId` to each group of $m \geq 2$ participants (m is a configuration parameter). Throughout a search session (which may include pre/post questionnaires), SearchX continuously captures fine-grained user activity logs such as clicks, hovers, and scrolls events. [1].

We now discuss the architecture of SearchX and then elaborate on three main design directions: supporting collaboration, empowering research, and retrieval providers and data collections.

2.3.1 Architecture Overview

When implementing SearchX, we chose to start from an existing system/interface to save development time. The options were limited as search engine interfaces are generally not open-sourced. We decided to use the single-user Pienapple search system [26][2] as a starting point, as it provides a generic Web search interface built with modern Web technologies (Node.js[3], React[4]) which has had up today an active developer communities and are supported by large companies, ensuring that the system will be relevant technology-wise for the upcoming years. Given this base system, we vastly expanded its functionalities for collaboration and experimentation, and then refactored the code base to be modular and reusable.

SearchX's client-server architecture is shown in Figure 2.1. The front-end is responsible for presenting the interface, managing task sessions, and logging user activities; the back-end is responsible for communicating with the retrieval engine, and managing group creation and synchronisation.

Front-end. The front-end (shown in Figure 2.3) is developed using React (a JavaScript library). React manages its own data model, minimising communication with the back-end; it enforces the creation of standalone view components, resulting in an interface simple to modify and extend. As the front-end is a Web application, any user with a modern browser can access it without requiring additional installation.

[1] A recent alternative for logging is using LogUI [89], a framework for capturing low-level logs in UI which are persistently stored in a data storage.

[2] The authors kindly provided us with their source code.

[3] https://nodejs.org/

[4] https://reactjs.org/

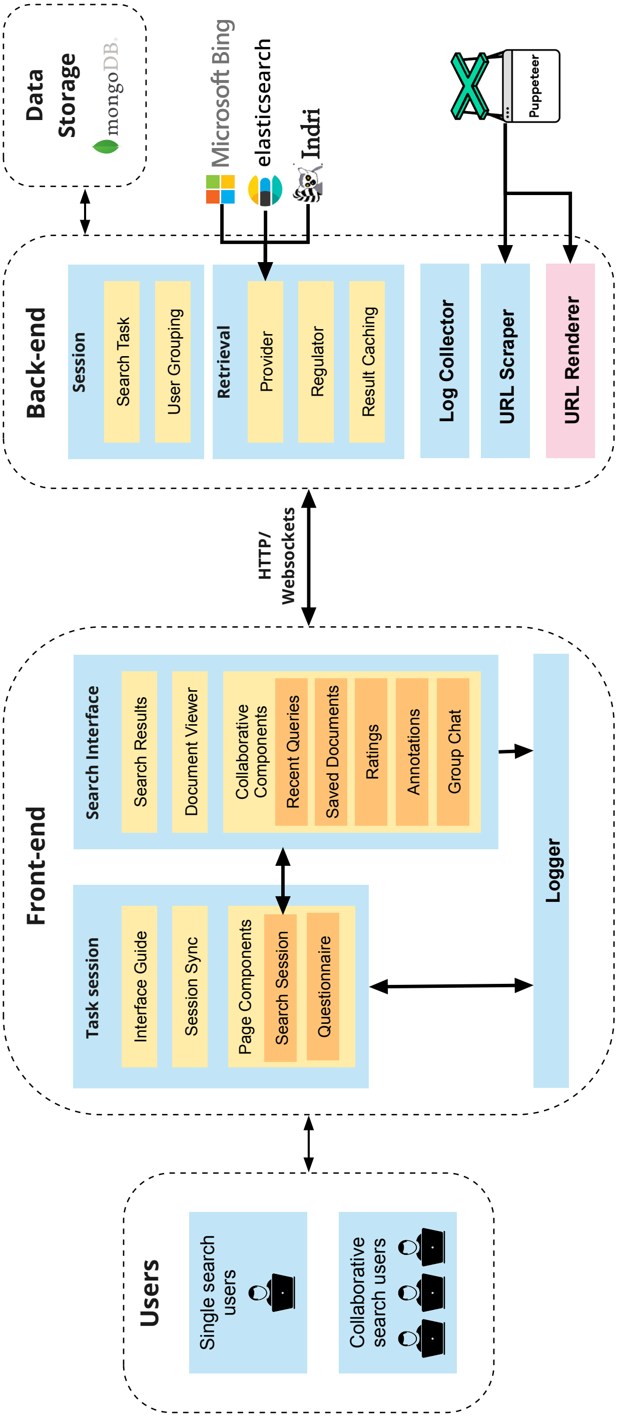


Figure 2.1: SearchX architecture overview.

The front-end consists of three logical abstractions. The *search interface* is composed of features related to searching and collaboration, and is presented to the user during the search session. Each feature is implemented as a standalone component which makes changing the layout or design of the interface efficient. Additionally, we separate the rendering component from the data management to make adjustments to the interface more efficient (e.g., adapting the interface for mobile or emerging devices). The *task session* component implements the desired experimental setup, and controls the search task, eventual group creation, and the experimental procedure (e.g., a pre-test, the search session, and then a post-test). An experimental procedure consists of a sequence of page routes, which we bootstrapped by implementing template components with one page route for the search session, and questionnaires, which we found to be the most commonly required templates in our experiments. The logger accepts activity data from each component, and regularly sends the logs to the back-end through an HTTP request for storage. This abstraction provides a clear separation of concern between interface features, experimental setup, and data collection, making it clear which part of the system needs to be changed for a particular experimental need.

Back-end. The back-end is developed with the `Node.js` server environment, which directly supports asynchronous I/O operations, making it suitable for applications requiring real-time updates. An added benefit of `Node.js` is its language (JavaScript)—developing both the front-end and back-end in the same language made the development more manageable for us. The back-end provides the application data services which are made available to the front-end through APIs—implemented using the `Express`[5] framework for HTTP and the `socket.io`[6] library for Web sockets. We chose these two libraries as they are currently the most common libraries for their respective role. We chose `MongoDB`[7] for the data storage as it uses a dynamic data schema, providing added flexibility during the development and modification of features.

The data services are categorised into four types. *Retrieval services* includes communication with the retrieval system through the provider, and further processing of the retrieval results through the regulator. Currently, we provide support for the Bing Search API for searching the Web, and support for `Elasticsearch` and `Indri`[8] servers for custom collections (see more details in Section 2.3.4). *Session services* handles group communication and assigning search tasks to users. *Collaboration services* includes the back-end logic of collaborative features in the front-end. *Utility services* includes data collection tools such as the log collector which stores user logs received from the front-end, and the URL scraper which scrapes all documents returned to the user. Additionally, we also have a URL renderer which makes it possible to load external Web pages inside our Web based system (think of a browser inside a browser), allowing us to implement the front-end document viewer. This offers the possibility to keep users inside the system at all times, allowing the system to log user interactions within the documents as well. Both the URL scraper and URL renderer utilise a headless browser via `Puppeteer`[9].

[5] https://expressjs.com/

[6] https://socket.io/

[7] https://www.mongodb.com/

[8] http://www.lemurproject.org/lemur/

[9] https://github.com/GoogleChrome/puppeteer

2.3.2 Supporting Collaboration

As can be seen in Table 2.1, only three prior systems implement system mediation, with each of them implementing a different type. In contrast, a number of interface mediation features (i.e. all features apart from system mediation) are popular across systems. SearchX also primarily supports collaboration through interface mediation, while facilitating custom implementation of system mediation through the addition of a regulator layer in the back-end. We have been developing SearchX for more than four years, and we have made four incremented versions over the four years. Figure 2.2 summarises the main changes we made in each version. Next, we discuss each implemented feature and its associated versions.

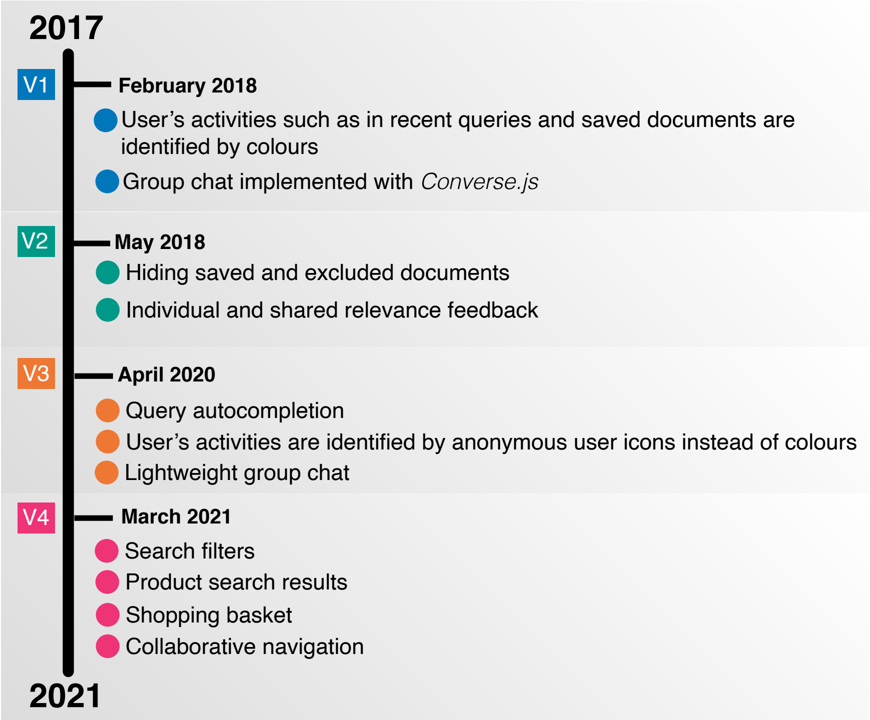
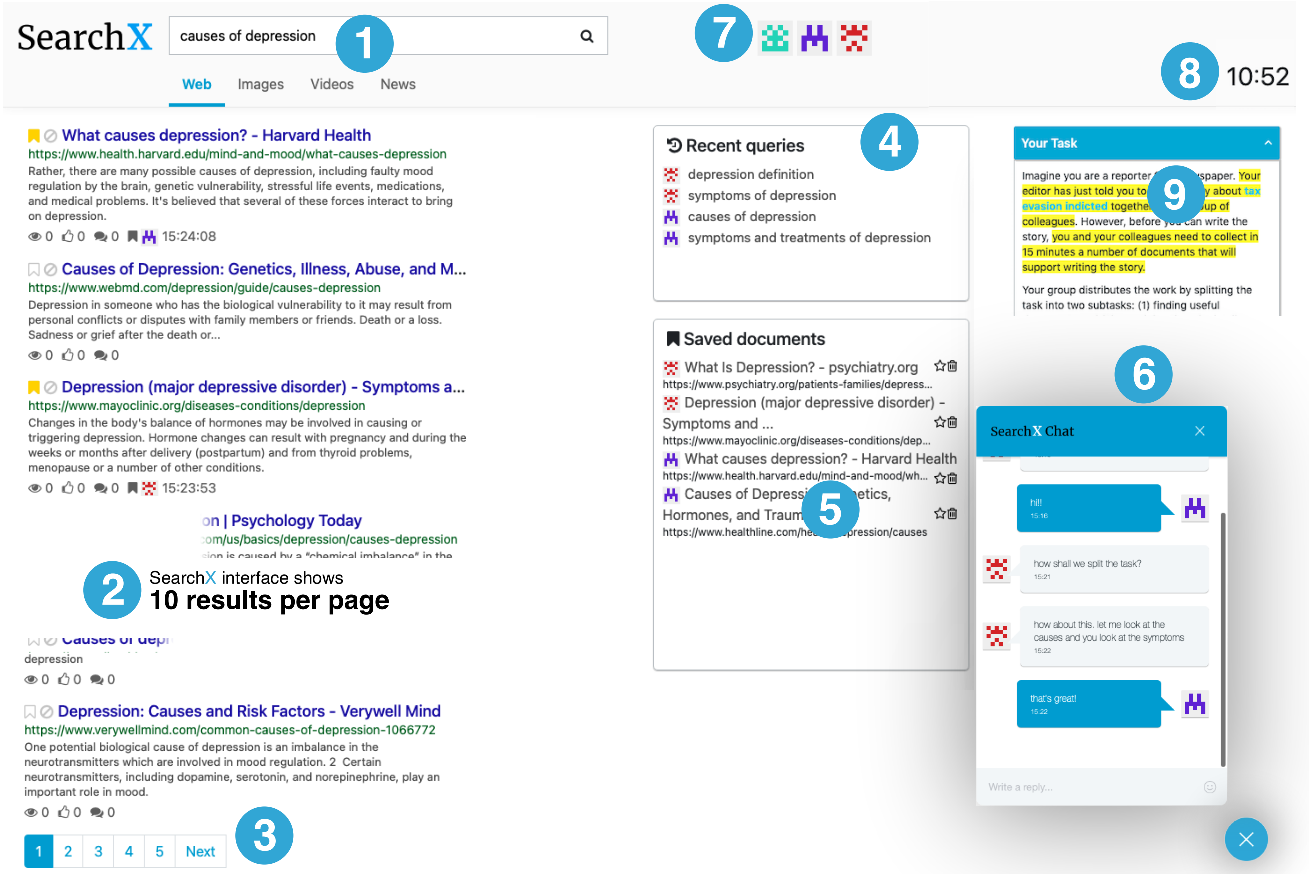

2017
V1
February 2018
User's activities such as in recent queries and saved documents are identified by colours
Group chat implemented with *Converse.js*
V2
May 2018
Hiding saved and excluded documents
Individual and shared relevance feedback
V3
April 2020
Query autocompletion
User's activities are identified by anonymous user icons instead of colours
Lightweight group chat
V4
March 2021
Search filters
Product search results
Shopping basket
Collaborative navigation
2021

Figure 2.2: Timeline of feature development of SearchX

Figure 2.3 shows the UI of our system with the basic features enabled using the Bing Search API as a retrieval provider. Complementary, Figure 2.7 shows the UI with additional components for conducting user studies with search results served by the Indri search system.

1 **Query Box** allows searchers to enter queries and select **search verticals**. For instance, commonly used Web search verticals are Web, Image, Video, or News, which can be provided using the Bing Search API provider.

Additionally, we provide with the query box, **query autocompletion (QAC)** in order to enable an user experience more in line with contemporary web search engines, we im-



SearchX

causes of depression

Web Images Videos News

10:52

What causes depression? - Harvard Health
https://www.health.harvard.edu/mind-and-mood/what-causes-depression
Rather, there are many possible causes of depression, including faulty mood regulation by the brain, genetic vulnerability, stressful life events, medications, and medical problems. It's believed that several of these forces interact to bring on depression.
0 0 0 15:24:08

Causes of Depression: Genetics, Illness, Abuse, and M...
https://www.webmd.com/depression/guide/causes-depression
Depression in someone who has the biological vulnerability to it may result from personal conflicts or disputes with family members or friends. Death or a loss. Sadness or grief after the death or...
0 0 0

Depression (major depressive disorder) - Symptoms a...
https://www.mayoclinic.org/diseases-conditions/depression
Changes in the body's balance of hormones may be involved in causing or triggering depression. Hormone changes can result with pregnancy and during the weeks or months after delivery (postpartum) and from thyroid problems, menopause or a number of other conditions.
0 0 0 15:23:53

on | Psychology Today
om/us/basics/depression/causes-depression

SearchX interface shows
10 results per page

depression
0 0 0

Depression: Causes and Risk Factors - Verywell Mind
https://www.verywellmind.com/common-causes-of-depression-1066772
One potential biological cause of depression is an imbalance in the neurotransmitters which are involved in mood regulation. 2 Certain neurotransmitters, including dopamine, serotonin, and norepinephrine, play an important role in mood.
0 0 0

1 2 3 4 5 Next

Recent queries
- depression definition
- symptoms of depression
- causes of depression
- symptoms and treatments of depression

Saved documents
- What Is Depression? - psychiatry.org
  https://www.psychiatry.org/patients-families/depress...
- Depression (major depressive disorder) - Symptoms and ...
  https://www.mayoclinic.org/diseases-conditions/dep...
- What causes depression? - Harvard Health
  https://www.health.harvard.edu/mind-and-mood/wh...
- Causes of Depress etics, Hormones, and Traum
  https://www.healthline.com/he pression/causes

Your Task

Imagine you are a reporter spaper. Your editor has just told you to y about tax evasion indicted togethe up of colleagues. However, before you can write the story, you and your colleagues need to collect in 15 minutes a number of documents that will support writing the story.

Your group distributes the work by splitting the task into two subtasks: (1) finding useful

SearchX Chat

hi!! 15:16

how shall we split the task? 15:21

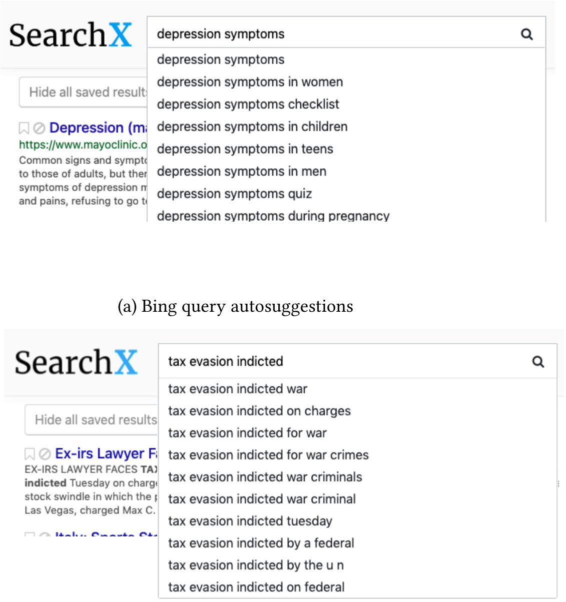
how about this. let me look at the causes and you look at the symptoms 15:22

that's great! 15:22

Write a reply...

Figure 2.3: The Basic SearchX interface, complete with latest style using Bing search results. Note that the chat interface is a popup—users can hide the popup to reveal all of the obscured interface components.

plemented the QAC functionality for the query box (released in version V3). Figure 2.4 shows an example of QAC using Bing Query Autosuggestions and collection-based query autocompletion. For the collection-based query autocompletion, inspired by the work of Bhatia et al. [20], we extracted all term uni-, bi-, tri-, and quad-grams from a document collection (e.g., TREC AQUAINT described in Section 2.3.4) as query autocompletion candidates and saved in a database with their frequency (e.g., `MongoDB`). For a given query, ten autocompletion candidates are selected by matching the query prefix to them. These are ranked by their frequency in given the document collection. For less common, *long-tail* QAC prefixes, a different approach was used inspired by Mitra and Craswell [92]. For example, assume no QAC candidates were available for the prefix `airport security`. The second term is taken in isolation (e.g., `security`), and all *n*-grams that begin with that term are then looked up in the database and ranked by their frequency. If no *n*-grams are found beginning with that term, the third term from the query is used, and so on. Suffixes are then extracted from these *n*-grams (i.e. `security checks`), with these suffixes then appended to the original query prefix, yielding a QAC candidate of `airport security checks`.



SearchX

depression symptoms

depression symptoms
depression symptoms in women
depression symptoms checklist
depression symptoms in children
depression symptoms in teens
depression symptoms in men
depression symptoms quiz
depression symptoms during pregnancy

(a) Bing query autosuggestions

SearchX

tax evasion indicted

tax evasion indicted war
tax evasion indicted on charges
tax evasion indicted for war
tax evasion indicted for war crimes
tax evasion indicted war criminals
tax evasion indicted war criminal
tax evasion indicted tuesday
tax evasion indicted by a federal
tax evasion indicted by the u n
tax evasion indicted on federal

Hide all saved results
Ex-irs Lawyer F
EX-IRS LAWYER FACES TAX
indicted Tuesday on charg
stock swindle in which the
Las Vegas, charged Max C.

(b) Collection-based query autocompletion

Figure 2.4: The SearchX interface search box with query autocompletion component. On the top the component is populated with Bing autosuggestions and on the bottom is populated with collection-based query autocompletion.

② **Search Results** are presented on the search result page (SERP) as *ten blue links*, each result complete with title, url, and snippet. Each result can either be saved by clicking 🔖 (which then appears in ⑤), or hidden from future SERPs by clicking ⊘. Clicking on a document title will show the full document within the interface and open a document viewer. Optionally, document metadata may be presented below each SERP entry to provide information about collaborators' activities on the document (e.g., number of views, annotations, ratings). This allows users to quickly identify documents that are considered relevant by their group. The information is presented using a number of simple icons.

③ **Search Results Pagination** are presented on the bottom of the SERP as *five buttons*. We removed the ability to jump to the last result as this may increase the workload of the retrieval provider.

④ **Recent Queries.** This feature has been implemented in all prior systems; it provides awareness of collaborators' search activities, allowing users to avoid duplication of effort and be inspired by their collaborators' choice of keywords [102]. Its implementation is similar across prior systems: as a list of queries that can be clicked on to immediately open results for that query. `SearchX` provides a scrollable list of recent queries in the sidebar ordered by time the query was issued.

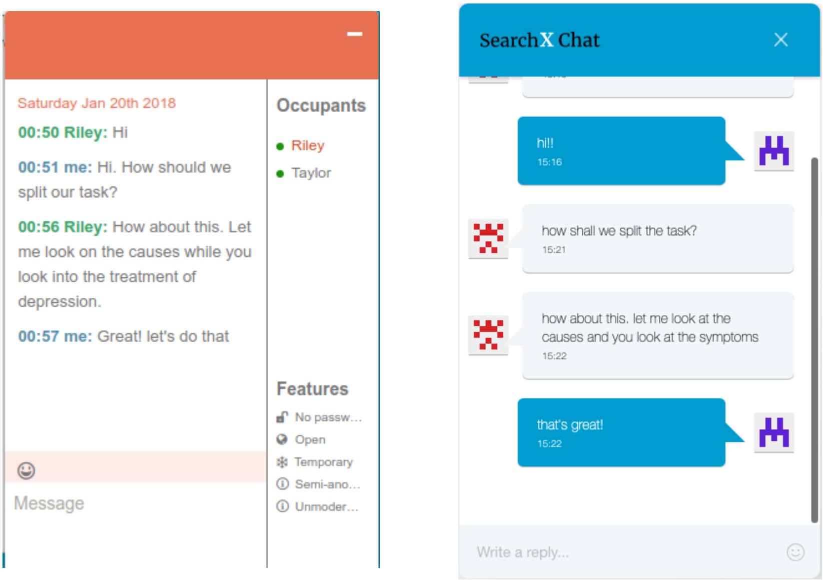
⑤ **Saved Documents.** Apart from functioning as a means to saved documents for later revisits, saving a document also promotes the shortlisting strategy which involves curating a shared list of potential documents [73]. Given the central role of shared saved documents in CSE research, we wanted to make it more accessible. Therefore we implemented the save button directly next to each search result. Furthermore, the list of saved documents is always visible in the sidebar to promote awareness of collaborators' actions; it is sorted by time, the most recently saved documents appear at the top. In addition, users benefit in the sense-making process when given the option to manage and rearrange their saved documents [73]. Therefore `SearchX` also implements pinned/starred saved documents which makes the document appear first in the saved documents list.

⑥ **Group Chat.** Even though knowledge sharing is already facilitated through more specialised mediation features, direct communication is necessary for coordination and discussions. We opted for a familiar pop-up design where the chat window is always visible in the interface but can be minimised when not in use (to avoid cluttering the interface). Up to `SearchX`'s version V2, we included group chat using `Converse.js`[10] which provides a robust chat window out of the box. A downside of `Converse.js` is that we could not have direct access to the internals, making it difficult to extend (e.g., we are not able to assign usernames automatically). Thus, in version V3, we developed our implementation `SearchX-chat` [11], which provide a lightweight chat `ReactJS` component, which can be easily plugged in and connect to a `SearchX` backend chat functionality using Websockets to send and receive messages instantly. We chose this option as we want the chat interface to employ anonymised user icons as detailed in component ⑦. Figure 2.5 shows a side-by-side example of the two chat versions. Notice that in our implementation, we opted to remove the names of the collaborators as a means to make them anonymous and neutral.

⑦ **User Status Bar and Users' Identification.** Up to `SearchX`'s version V2, we colour coded elements of the interface that are associated with a particular collaborator's actions (such as querying and saving documents). This allows users within the group to differen-

[10] https://conversejs.org/

[11] https://github.com/searchx-framework/searchx-chat



Saturday Jan 20th 2018

00:50 Riley: Hi

00:51 me: Hi. How should we split our task?

00:56 Riley: How about this. Let me look on the causes while you look into the treatment of depression.

00:57 me: Great! let's do that

hi!! 15:16

how shall we split the task? 15:21

how about this. let me look at the causes and you look at the symptoms 15:22

that's great! 15:22

(a) `Converse.js` group chat interface

(b) Our group chat interface

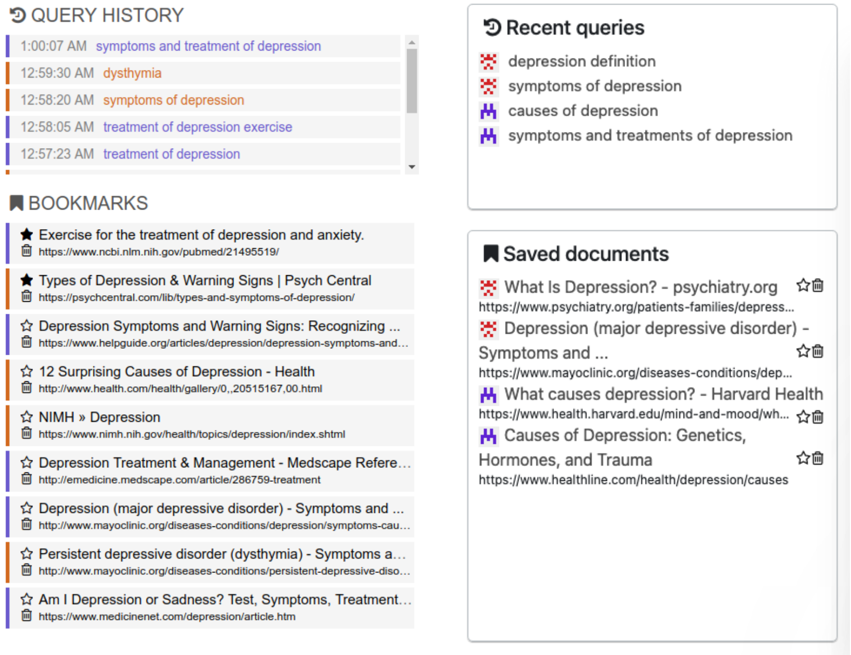
Figure 2.5: The `SearchX` chat component. On the left the component implemented `Converse.js` colour coded and on the right with our own implementation.

tiate between the activities and contributions of each individual collaborator. The colours are generated randomly. However, from version V3, we chose to implement a coloured *Identicon* to provide a means for identifying who did what using an approach often employed by contemporary websites offering collaborative functionalities (such as *GitHub*, or *Google's* collaborative office tools, which use *anonymous animals*). These icons were present in the status bar (⑦). If a user hovers over each icon, a popup with the user names or anonymized IDs is shown. Also, user icons are placed next to the queries issued by a given user in ④, with the same principle employed in component ⑤. Figure 2.6 shows a comparison among the two versions. These icons are also present on the chat interface ⑥. The inclusion of this feature permits users to identify their collaborators in an anonymized fashion, instead of using their own names—which could introduce biases (e.g., through gender identification, nationality).

⑧ **Timed Session.** A **countdown timer** is present at the top-right of the interface, counting down to 0:00. The countdown clock is synchronised across participants' systems.

⑨ **Task Bar.** The task outline is provided for participants to examine whenever they wish during the CSE task.

System Mediation. As stated before, `SearchX` does not implement a specific form of system mediation, but facilitates such an implementation if needed. In Section 2.2 we outlined that system mediation is usually performed in the form of modifications to the retrieved list of results (i.e., re-ranking). We have designed the retrieval service in the back-end to also contain a *regulator layer* that enables us to adjust the SERP sent to each



QUERY HISTORY

1:00:07 AM symptoms and treatment of depression
12:59:30 AM dysthymia
12:58:20 AM symptoms of depression
12:58:05 AM treatment of depression exercise
12:57:23 AM treatment of depression

BOOKMARKS

★ Exercise for the treatment of depression and anxiety.
https://www.ncbi.nlm.nih.gov/pubmed/21495519/

★ Types of Depression & Warning Signs | Psych Central
https://psychcentral.com/lib/types-and-symptoms-of-depression/

☆ Depression Symptoms and Warning Signs: Recognizing ...
https://www.helpguide.org/articles/depression/depression-symptoms-and…

☆ 12 Surprising Causes of Depression - Health
http://www.health.com/health/gallery/0,,20515167,00.html

☆ NIMH » Depression
https://www.nimh.nih.gov/health/topics/depression/index.shtml

☆ Depression Treatment & Management - Medscape Refere…
http://emedicine.medscape.com/article/286759-treatment

☆ Depression (major depressive disorder) - Symptoms and ...
http://www.mayoclinic.org/diseases-conditions/depression/symptoms-cau…

☆ Persistent depressive disorder (dysthymia) - Symptoms a…
http://www.mayoclinic.org/diseases-conditions/persistent-depressive-diso…

☆ Am I Depression or Sadness? Test, Symptoms, Treatment…
https://www.medicinenet.com/depression/article.htm

(a) Colour code for each user

Recent queries

depression definition
symptoms of depression
causes of depression
symptoms and treatments of depression

Saved documents

What Is Depression? - psychiatry.org ☆
https://www.psychiatry.org/patients-families/depress...

Depression (major depressive disorder) - Symptoms and ... ☆
https://www.mayoclinic.org/diseases-conditions/dep...

What causes depression? - Harvard Health
https://www.health.harvard.edu/mind-and-mood/wh... ☆

Causes of Depression: Genetics, Hormones, and Trauma ☆
https://www.healthline.com/health/depression/causes

(b) User icons

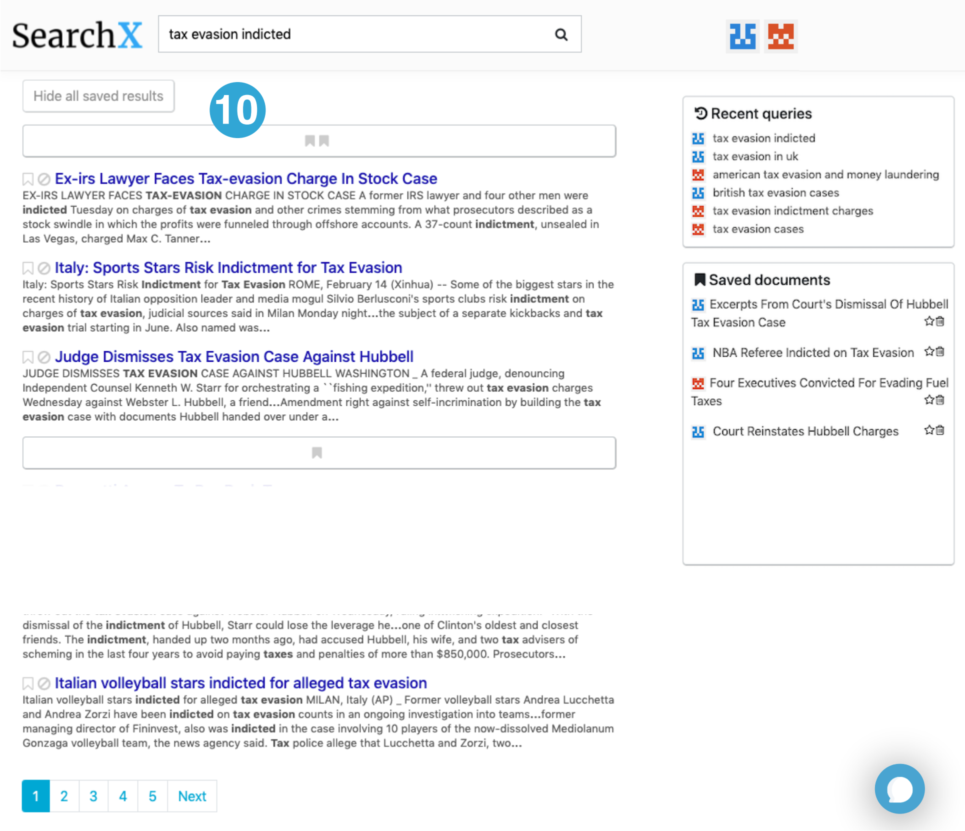
Figure 2.6: The SearchX collaborative components of recent queries and saved documents. On the left the components use the own style with colour coded and on the right with user icons generated with `Identicon.js`.

collaborator based on the actions of the group's members.

The regulator layer collects the necessary input data for system mediation by fetching and aggregating it from the MongoDB database. One example of such data is the current collection of saved documents and up-voted results for the entire group, which can be used as input for relevance feedback. The input data can be sent to the search provider in order to incorporate the data in the retrieval algorithm. This is useful to incorporate features from the search provider into system mediation. For example, Indri supports relevance feedback by sending it a set of results. The input data can also be used directly in the regulator to re-rank or filter the list of results. This option can be used for distribution of labour by filtering the documents that are assigned to each user according to a distribution criterion. Figure 2.7 shows an example of SearchX with division of labour by the **Hiding Saved Results** buttons 10, which was implemented in version V2. Once a user saves a document, for the next query, collaborators saved search results are automatically hidden.

Document Viewer. When the user clicks on a document link (via 2 or 5), the document viewer is shown. In Figure 2.8 we show the document viewer with the following components.

11 **Document Rating.** Document rating is mainly considered as fine-grained source of information for relevance feedback. To avoid cluttering of the SERP, we present the rating



Ex-irs Lawyer Faces Tax-evasion Charge In Stock Case
EX-IRS LAWYER FACES **TAX-EVASION** CHARGE IN STOCK CASE A former IRS lawyer and four other men were **indicted** Tuesday on charges of **tax evasion** and other crimes stemming from what prosecutors described as a stock swindle in which the profits were funneled through offshore accounts. A 37-count **indictment**, unsealed in Las Vegas, charged Max C. Tanner...

Italy: Sports Stars Risk Indictment for Tax Evasion
Italy: Sports Stars Risk **Indictment** for **Tax Evasion** ROME, February 14 (Xinhua) -- Some of the biggest stars in the recent history of Italian opposition leader and media mogul Silvio Berlusconi's sports clubs risk **indictment** on charges of **tax evasion**, judicial sources said in Milan Monday night...the subject of a separate kickbacks and **tax evasion** trial starting in June. Also named was...

Judge Dismisses Tax Evasion Case Against Hubbell
JUDGE DISMISSES **TAX EVASION** CASE AGAINST HUBBELL WASHINGTON _ A federal judge, denouncing Independent Counsel Kenneth W. Starr for orchestrating a ``fishing expedition,'' threw out **tax evasion** charges Wednesday against Webster L. Hubbell, a friend...Amendment right against self-incrimination by building the **tax evasion** case with documents Hubbell handed over under a...

dismissal of the **indictment** of Hubbell, Starr could lose the leverage he...one of Clinton's oldest and closest friends. The **indictment**, handed up two months ago, had accused Hubbell, his wife, and two **tax** advisers of scheming in the last four years to avoid paying **taxes** and penalties of more than $850,000. Prosecutors...

Italian volleyball stars indicted for alleged tax evasion
Italian volleyball stars **indicted** for alleged **tax evasion** MILAN, Italy (AP) _ Former volleyball stars Andrea Lucchetta and Andrea Zorzi have been **indicted** on **tax evasion** counts in an ongoing investigation into teams...former managing director of Fininvest, also was **indicted** in the case involving 10 players of the now-dissolved Mediolanum Gonzaga volleyball team, the news agency said. **Tax** police allege that Lucchetta and Zorzi, two...

Recent queries
- tax evasion indicted
- tax evasion in uk
- american tax evasion and money laundering
- british tax evasion cases
- tax evasion indictment charges
- tax evasion cases

Saved documents
- Excerpts From Court's Dismissal Of Hubbell Tax Evasion Case
- NBA Referee Indicted on Tax Evasion
- Four Executives Convicted For Evading Fuel Taxes
- Court Reinstates Hubbell Charges

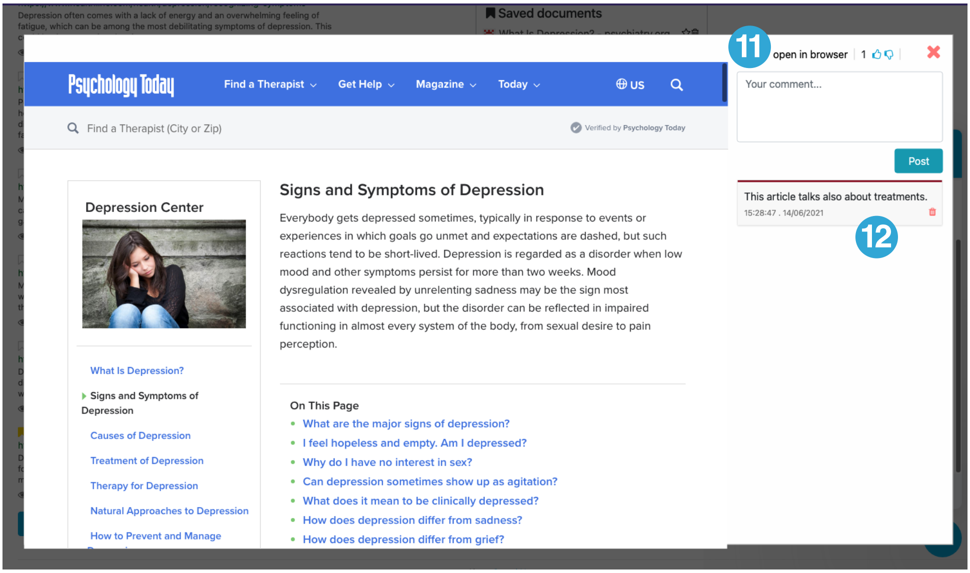
Figure 2.7: The SearchX interface with hiding of saved results, and task bar with the latest style using Indri search results (TREC AQUAINT collection). Note that the chat interface button, click on it and it reveal the chat component.

buttons not on the SERP, but inside the document viewer; the added benefit is that users can only rate once they have seen the document. Document rating is implemented as a like/dislike button to leverage users' familiarity with this type of interaction.

12 **Document Annotation.** Unlike existing systems, we implemented annotations as a message thread similar to chat interfaces. This setup highlights the bidirectional nature of the annotation process, promoting sense-making through the exchange of opinions. The annotation interface is presented inside the document viewer, directly next to the document to make adding new annotations a quick process.

Lastly, in the most recent version of SearchX, version V4, we extended SearchX with E-commerce features with the following components as shown in Figure 2.9.

13 **Search Filters**. We provide users with the ability to filter by store department (e.g., Electronics, Kitchen, and Home), average customer rating, brand, and price. Once someone clicks in one of the filters, the server returns filtered search results. A reset button is



Depression often comes with a lack of energy and an overwhelming feeling of fatigue, which can be among the most debilitating symptoms of depression. This

This article talks also about treatments.
15:28:47 . 14/06/2021

Signs and Symptoms of Depression

Everybody gets depressed sometimes, typically in response to events or experiences in which goals go unmet and expectations are dashed, but such reactions tend to be short-lived. Depression is regarded as a disorder when low mood and other symptoms persist for more than two weeks. Mood dysregulation revealed by unrelenting sadness may be the sign most associated with depression, but the disorder can be reflected in impaired functioning in almost every system of the body, from sexual desire to pain perception.

On This Page

- What are the major signs of depression?
- I feel hopeless and empty. Am I depressed?
- Why do I have no interest in sex?
- Can depression sometimes show up as agitation?
- What does it mean to be clinically depressed?
- How does depression differ from sadness?
- How does depression differ from grief?

Figure 2.8: Document viewer

implemented to remove all currently enabled filters for the issued query. Search filters are not available in standard search. One must implement one for each data collection and adapt the search filters.

14 **Product SERP**. Product search results are presented on the SERP as 12 cards (three per row); complete with the main product image, product title, average rating, number of reviews, and product price. Pagination is also present, with links provided at the bottom of the results (not shown in Figure 2.3; results have been cropped to save space).

15 **Product Card Result**. For each result, users can *save* the item by clicking the ♥ icon. We chose for a ♥ icon as this is more contemporary in e-commerce interfaces. If users wish to take an item forward, they can click the 🛒 icon.

16 **Alerts**. In case a shared navigation mechanism is enabled (see Chapter 6 for more details), when another user issues a new query, changes verticals or filters results, this popup box will appear to warn the other group members. After three seconds, the interface will be updated to reflect the new set of (filtered) results.

17 **Search Facets**. Similarly to search verticals, users can select from different *facets* or categories of items pertaining to results presented to them.

18. **Task Outline Button**. To allow more space for the product cards, a *task outline* button is provided for users to show the task whenever they wish to view it.

19 **Saved Items** We modified the *Saved Documents* widget, to enable a 🛒 button. The *saved items* component lists each of the items that were saved by the users; see 13.

20 **Shopping Basket**. This component lists the items that users have decided to take forward in their shopping.

2.3.3 Empowering Research

We now elaborate on how SearchX was designed as a tool for research.

Availability and Accessibility. SearchX is open-sourced for use and development by other researchers. We iterated on the installation process a number of times to make it simple and effective. We provide three example implementations of different experimental setups (synchronous CSE, asynchronous CSE, single-user search) to be expanded upon. We also put significant effort into extensively documenting how researchers can modify the system (e.g., by adding new UI features, or by changing the retrieval system in the back-end), and this has helped other researchers to take up SearchX for their research [28, 119, 121, 123].

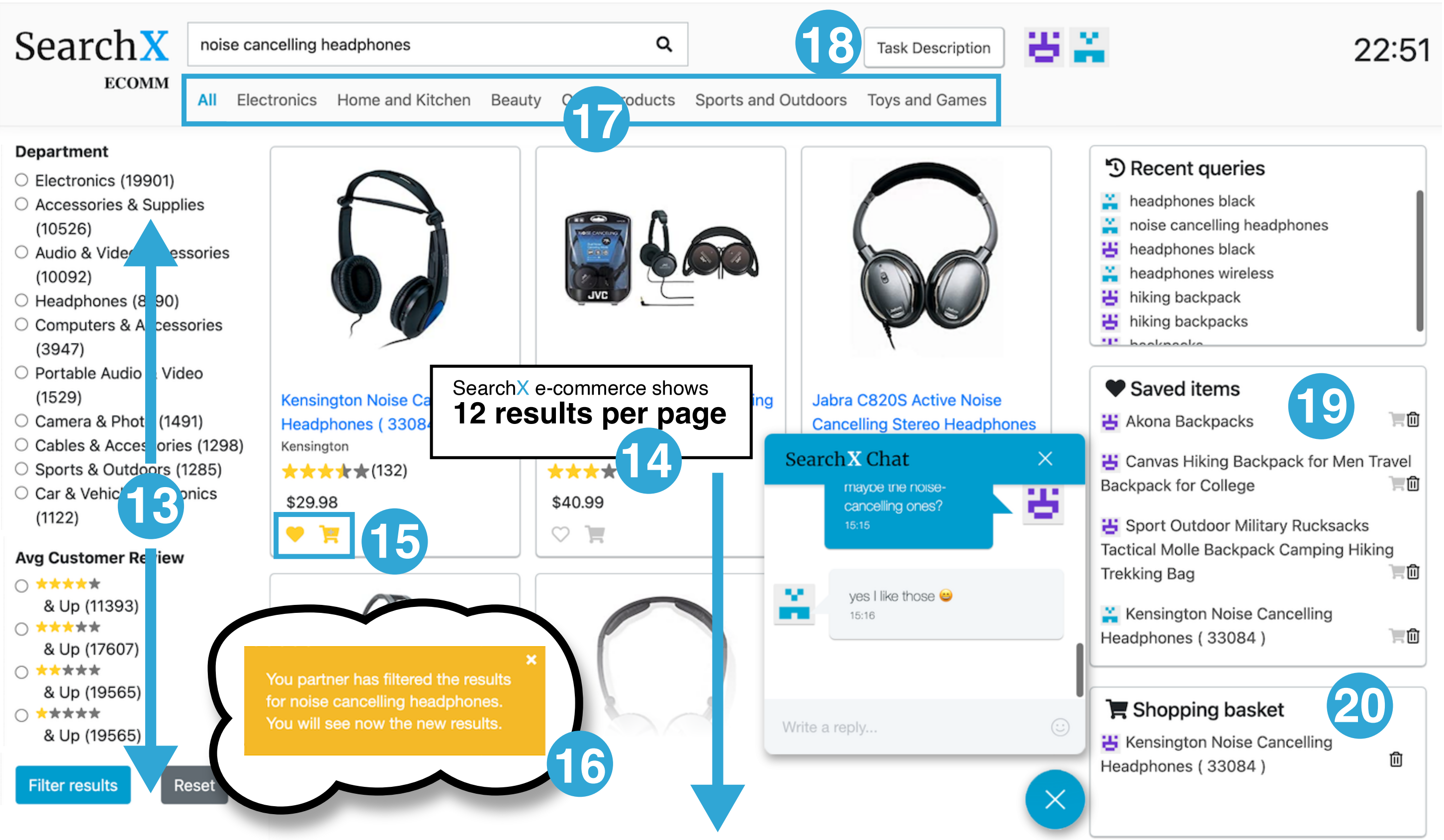
Study Creation. Currently, modifying the system requires programming knowledge as we do not provide a graphical interface to create user studies yet. However, we have created reusable implementations of common components in the experimental setup: questionnaires and the search session. The questionnaires are implemented using SurveyJS[12], which allows defining questionnaires directly using JSON. We have created a React component that abstracts over SurveyJS, adding logging features and flow control. We also did the same for the search session, which abstracts over the search interface, adding session-related logs, flow control, and a task bar to describing the search task. We found this to simplify the creation of new user studies, since it takes away much of the boilerplate code needed in configuring the experimental procedure.

Data Collection. A requirement for a CSE user study is the collection of user activity logs. In SearchX, we have added logging to all interactive components of the system so that it records when a user hovers over or directly interacts with a component (e.g., clicking, querying, opening a document). We also log session related data (e.g., starting/finishing the search session, submitting a questionnaire) and interactions with the browser (e.g., changing tabs), which helps understanding all actions executed by a user. All logs are captured directly by the interface without the need for third party plugins installed by the user. All logs are defined and implemented in the front-end, while the back-end only handles storage of logs, making it easy to modify the logs or create additional logs.

Interface Guide. Prior works report that some features of their system were not explored much by users because they do not know or understand it [54, 102]. We solve this issue by adding a guided interface walk-through of the interface (built using Intro.js[13]) which explains step-by-step what each feature is meant to do. Figure 2.10 shows one step example of the interactive guide. This interface guide is launched when a user first starts the search session, ensuring that they are aware of the features we want them to use, before moving on to the search task.

[12] https://surveyjs.io/

[13] https://introjs.com/



noise cancelling headphones

SearchX e-commerce shows 12 results per page

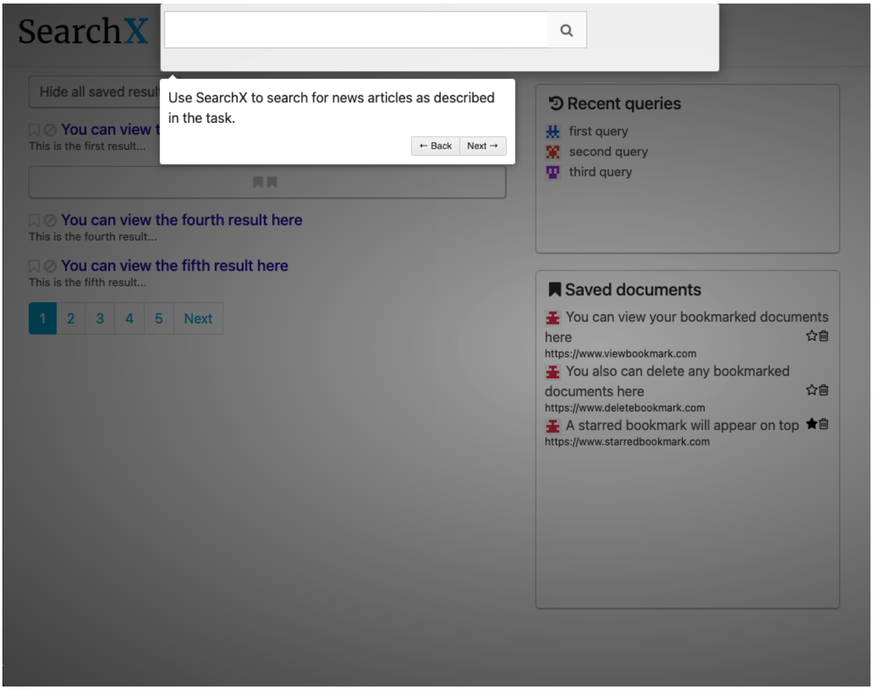
You partner has filtered the results for noise cancelling headphones. You will see now the new results.

maybe the noise-cancelling ones? 15:15

yes I like those 😀 15:16

SearchX
ECOMM
Task Description
22:51
All
Electronics
Home and Kitchen
Beauty
Sports and Outdoors
Toys and Games
Department
○ Electronics (19901)
○ Accessories & Supplies (10526)
○ Audio & Video Accessories (10092)
○ Headphones (8890)
○ Computers & Accessories (3947)
○ Portable Audio & Video (1529)
○ Camera & Photo (1491)
○ Cables & Accessories (1298)
○ Sports & Outdoors (1285)
○ Car & Vehicle Electronics (1122)
Avg Customer Review
○ ★★★★★ & Up (11393)
○ ★★★★★ & Up (17607)
○ ★★★★★ & Up (19565)
○ ★★★★★ & Up (19565)
Filter results
Reset
Kensington Noise Cancelling Headphones (33084)
Kensington
★★★★★ (132)
$29.98
$40.99
Jabra C820S Active Noise Cancelling Stereo Headphones
SearchX Chat
Write a reply...
Recent queries
headphones black
noise cancelling headphones
headphones black
headphones wireless
hiking backpack
hiking backpacks
backpacks
Saved items
Akona Backpacks
Canvas Hiking Backpack for Men Travel Backpack for College
Sport Outdoor Military Rucksacks Tactical Molle Backpack Camping Hiking Trekking Bag
Kensington Noise Cancelling Headphones (33084)
Shopping basket
Kensington Noise Cancelling Headphones (33084)
13
14
15
16
17
18
19
20

Figure 2.9: The SearchX with components adapted for e-commerce. Search results are returned using Elasticsearch indexed with Amazon collection. Note that the chat interface is a popup—users can hide the popup to reveal all of the obscured interface components.



Use SearchX to search for news articles as described in the task.

first query
second query
third query

You can view t
This is the first result...

You can view the fourth result here
This is the fourth result...

You can view the fifth result here
This is the fifth result...

You can view your bookmarked documents here
https://www.viewbookmark.com

You also can delete any bookmarked documents here
https://www.deletebookmark.com

A starred bookmark will appear on top
https://www.starredbookmark.com

Figure 2.10: The `SearchX` interactive introduction guide powered by `Intro.js`

2.3.4 Retrieval Providers and Data Collections

Now we describe two retrieval providers, `Indri` and `Elasticsearch`, the collections we use, and how we index and retrieve the results for these collections.

Indri

The IR field is aided by numerous efficient search engine implementations, aimed at research and industry, such as `Indri` [139], `Lucene`[14], `Terrier` [108] and `Anserini` [164]. In this thesis, we make `Indri` accessible to the modern web stack. Many modern web applications and frameworks make use of `Node.js`. A significant advantage of this framework is the single programming language on the client and server-side (JavaScript), which simplifies development; in addition, `Node.js` is highly scalable [149]. In order to design and evaluate web search interfaces, a backend, implemented in `Node.js`, requires access to a search system. One option is to call `Indri` via system calls. However, the disadvantage of system calls via shell commands is the extra layer of communication with the operating system.

We design an alternative, `node-indri`, a `Node.js` module implemented with an easy-to-use API. It provides access to basic `Indri` functionalities such as search with relevance feedback and document scoring. Importantly, `node-indri` is implemented in a non-blocking manner. `node-indri`'s development started with the need to make `Indri`'s state-of-the-

[14] `http://lucene.apache.org/`

art relevance feedback models accessible to students, that (i) tend to have little experience with C++, but are familiar with modern web programming paradigms and (ii) are not IR experts and thus struggle to make sense of `Indri`'s internals.

In Table 2.2, we list the arguments of `node-indri`'s three classes. Each class has at most two methods with arguments that depend on the functionalities exposed from `Indri`. The last argument is a callback function implementing the error-first pattern. In this manner, `node-indri` is an asynchronous module, with most of these functions assessing lower-level system functionalities through `libuvl`. This in turn means that the methods are executed in `Node.js`' thread pool, making `node-indri` naturally parallel.

Table 2.2: Overview of the arguments necessary for `node-indri`'s method calls. Our API is simple and includes only one method per class. The last argument is always a callback that is executed when the data has been retrieved. Underlined are the required parameters.

`Searcher.search`	`Reader.getDocument`	`Scorer.scoreDocuments`	`Scorer.retrieveTopKScores`
query, page, results-per-page, feedback-docs, callback	docid, callback	query, docs, callback	query, number-of-results, callback

The models' hyperparameter settings (e.g., μ in the case of language mode ling with Dirichlet smoothing) are manually set via a configuration file. We now discuss the goal of each of the three classes `node-indri` makes available to its users in turn:

`Searcher` This class exposes the functionalities of `Indri`'s `QueryEnvironment` and `RMExpander` classes through the method `search` which returns a list of search results in a paginated manner. When a `Searcher` object is instantiated, it takes a configuration object as argument (these settings include the retrieval models' hyperparameters and flags of the type of data to return). When a call to `search()` is made and no feedback documents are provided as argument, the standard query likelihood model is employed, otherwise RM3 model is [83]. Depending on the configuration settings, the returned result list may contain document snippets (as provided by `Indri`'s `SnippetBuilder`), document scores, document text and other metadata.

`Reader` This class exposes the functionalities of an `Indri` index through the method `getDocument` in order to return a document's meta- and text data.

`Scorer` This class provides access to the retrieval scores of a list of documents via the method `scoreDocuments`. In addition, it provides `retrieveTopKScores` to retrieve the scores and document ids of the top ranked documents for a query.

Collection. We use the TREC AQUAINT corpus for the experiments in Chapter 3 and Chapter 4. The corpus contains a total of 1,033,461 documents (or newspaper articles) from three separate newswires: the *New York Times (NYT)*; the *Associated Press*; and *Xinhua (XIE)*. The articles included within the corpus were collected during the period 1996-2000. TREC AQUAINT was also the collection of choice in the TREC 2005 Robust track [155].

We create a TREC AQUAINT index using the corpus with stopword removal with *Indri* stopwords, and a Krovetz stemmer applied. The *Indri* retrieval system was used. Before

Table 2.3: Data Collection statistics including the filtered products considered in this work.

Product Department	**Used in this thesis**	**Provided in [91]**
Electronics	364,099	498,196
Home and Kitchen	288,880	436,988
Beauty	176,670	259,204
Office Products	122,406	134,838
Sports and Outdoors	244,366	532,197
Toys and Games	230,858	336,072
All departments	**1,346,033**	**2,197,495**

indexing, a document was checked to see whether: (i) it had a title; and (ii) a document with the same title had been indexed previously. These checks were considered to ensure that all search results from the completed index looked reasonable, and that duplicates were not present, nor documents with `Untitled` as the document title.

For (ii), near duplicates were detected using *SimHash* with parameters $blocks = 4$ and $distance = 3$, following the work by [87]. Intuitively, $distance$ is the bumber of bits that may differ in matching pairs of documents, and $blocks$, the number of blocks used that divide a hashed string of each document. This returned a set of documents judged by the method to be duplicates, and from that set of documents, we opted to keep the document with a relevant judgement in the TREC 2005 Robust, while dropping the remaining documents. If multiple documents judged to be relevant were returned, we randomly considered a single document—leading to a revised number of relevant documents in the index (1187 from the 6479 (18.32%) relevant documents were deemed duplicated by *SimHash*). Overall, a total of 854,130 documents (82.65%) remained once these two cleaning steps had been completed.

During the experiments, we also relied on Indri for snippet generation. Query-biased snippets were extracted from associated documents for presentation within each rendered SERP. Figure 2.7 shows an example of query-biased generated snippets for the query “tax evasion indicted”. Results were returned using language modelling (LM) with Dirichlet smoothing [168] with hyperparameter μ set to 2500.

Elasticsearch

`Elasticsearch` is a distributed, free and open search and analytics engine built on top of Apache Lucene and was first released in 2010. `Elasticsearch` for all types of data, including textual, numerical, geospatial, structured, and unstructured. In Chapter 6, we create an index of an e-commerce collection using `Elasticsearch` as our retrieval provider. We chose `Elasticsearch` because it provides the functionality to the search filters presented in 11.

Collection. We utilise the Amazon datasets collected by McAuley et al. [91], which has also been widely employed in previous product search research [21, 57, 150, 153, 154, 171]. We selected six different product domains (Amazon departments) for our study as listed in Table 2.3. We chose these categories in order to provide a diverse product catalogue. In addition, these categories are common in the product search literature [21, 57, 153, 154, 171]. We filtered out any product without a title, description, price, or image, as these are essential for the UI. In total, 1,346,033 products were indexed; 851,462 products were skipped. In addition to the aforementioned product information, we also extracted the product's department, reviews, average rating (and thus could show the top five reviews in the product viewer, as shown at ⑦ in Figure 2.3). Each review consists of a rating, the review text, and the number of votes cast for the review being helpful.

Indexing. Each product field is tokenised using character n-gram tokeniser ($min = 2$, and $max = 10$), which first breaks text down into words whenever it finds one of a list of specified characters, then it produces N-grams of each word where the start of the N-gram is anchored to the beginning of the word (e.g., *The* becomes *T, Th, The*). Furthermore, in order to use the values in the numeric fields to rank search results, we indexed the fields average rating, the number of reviews, and average sales rank as `rank_feature` field type. This field type allows us to store real values for later be used as part of a customised ranking function, which otherwise would not be possible.

Retrieval. For this dataset, results were returned using `rank_feature` query [15] with BM25 as base retrieval function with default parameters. We employed a multi-field retrieval based on the product title, description, reviews, departments, and brand as text fields, where the best score of all the fields is used. As we did not have a ground truth dataset, we selected a few hyperparameters by manually inspecting the search results returned by queries that included a product type together with one attribute (e.g., black headphones, white desk, German jersey). For field weights, we set product title field weight two and the rest field weight one. For the `rank_feature` query weights, we set weights for the fields average rating, number of reviews, and average sales rank to 50, 2, and 1, respectively.

2.4 Challenges and Limitations

We now discuss issues that are usually hidden from view—things that did not go as intended or slowed down the process.

2.4.1 Iterating on the Experimental Setup

We initially implemented the basic version of `SearchX` with a paper deadline in mind. This led to a working but not very modular version of `SearchX`, which we realised when attempting to implement a number of CSE experiments—for each experiment, multiple files in both the front-end and back-end required changes. Since we wanted the system to be reusable for different experiments, we invested effort onto refactoring the code for a more intuitive experimental setup. We started fixing this in the front-end by separating out all code related to the experimental setup from the search interface and encapsulating them into reusable React components. While this simplified the experimental setup, the

[15] `rank_feature` query calculates relevance scores based on rank feature fields. See more details on `https://www.elastic.co/guide/en/elasticsearch/reference/7.14/query-dsl-rank-feature-query.html`

communication with the back-end remained a complex issue. We now limit the responsibility of the back-end to only group management and synchronisation, allowing us to directly implement the limited range of functionality inside the task components. If we would have spent more time on the initial design, we would have saved substantial development time, iterating a number of times on the architecture and the interactions among the components.

2.4.2 Deploying a Crowdsourced Study

As an effort to support online studies, we adapted SearchX for crowd-sourced studies. During a first CSE pilot on FigureEight (formerly CrowdFlower), we found crowd workers to not be overly motivated to properly execute our assigned CSE tasks (many tasks on FigureEight tend to be short and do not require elaborate instructions such as image labelling). We found two ways around this issue: (i) a new platform and (ii) actively encouraging complying behaviour. We switched to the research-focused Prolific[16] platform which was shown to provide higher quality data [110]—something we found to be true as well in our work. We also spent significant development time on monitoring workers' attentiveness and actively keeping them on track. We logged browser interactions (change tabs, context menu) and notified workers about their tab changes in real-time (after *n* tab changes a worker is no longer paid). We also added quality control questions and disabled copy and paste operations in the questionnaires. Complying behaviour is a common issue in crowdsourcing user studies, especially long user studies as we conduct along this thesis. A body of work on suggestions on how to insure high qualify of the data was consulted during this thesis [2]. All these steps improved the quality of data we collected, but were slow to be implemented as we discovered them as solutions to worker compliance issues one by one after running another (and another) pilot study.

2.4.3 Synchronising Collaborative Sessions

Running a synchronous search session through a crowdsourcing platform is tricky, since workers are not available right away, therefore a type of "waiting room" is needed for the grouping so that workers assigned to a single group start their search session at the same time. This problem becomes particularly intense as the group size increases—an experiment with 20 workers requires 20 workers to accept the task at roughly the same time. Another issue we encountered was that workers were disconnected from the grouping process when the page was refreshed/closed during the waiting period, resulting in the worker not being able to continue the study. We currently just warn workers that attempt to refresh/leave the Web page running SearchX but a better way to handle (and entertain?) workers in the "waiting room" is needed to enable CSE experiments with large group sizes. Additionally, we also added a Snake game and a Tetris game as entertainment for groups in the "waiting room".

2.4.4 Implementing a Document Viewer.

Ensuring that crowdworkers remain within SearchX (and otherwise rescind the payment) is a good way of ensuring compliance, but of course this idea breaks down when we want

[16] https://prolific.ac/

the workers to interact with the SERP (and click on links and view documents in another browser tab). We thus needed to implement a document viewer (again requiring valuable development time) that allows users to view the document within SearchX. This is rather straightforward for static resources such as text or images, however it is not possible to render another Web page directly inside SearchX because of CORS (cross origin resource sharing) restrictions. We thus had to render the URL in the back-end and pass the rendered HTML to an iframe in the front-end. This is an imperfect solution though since the resulting page is static with most interactive elements disabled, and at times the rendering is not perfect. We are still improving this aspect of SearchX. Currently, to alleviate the issue of imperfect renders, we add a button to open the web page in another tab as shown in Figure 2.8.

2.4.5 Dynamic Result List and System Mediation

If the regulator layer of SearchX is used, the list of results is no longer a function of only the user's query, because it can change based on other input data as described in Section 2.3.2. In our experience this can lead to several challenges.

For instance, one aspect that needs to be considered when implementing system mediation is the interaction of the mediation features with the search interface. A result list that is modified can lead to a jarring user experience, especially if the list updates in real time. We consider that it is better to apply changes to the SERP after a user initiates an action (e.g., a new query or page change); by combining the update due to system mediation with a user-initiated update of the SERP the user is not confused that the page changes. Another approach to prevent confusion is to indicate to users when results have been omitted or re-ranked and to give them the option to enable and disable mediation features. In this manner, users are given the autonomy to decide when system mediation features are useful.

2.4.6 Synchronised Algorithmic Mediation

Shared relevance feedback and division of labour can be implemented in one of two ways: either immediately or delayed. In the immediate version, as soon as a collaborator saves a document that action should be reflected in the SERPs of all collaborators—not just by updating the shared saved documents list and hiding the document in question from the SERPs, but also by rerunning each user's submitted query with the new set of relevant documents. This is likely to confuse users as they cannot anticipate when (and why) a result ranking suddenly changes; worse still, if a user paginates through the result list, she might miss the newly highly ranked relevant documents because she is looking at lower ranks. We overcome this issue by opting for delayed relevance feedback and division of labour: only when a collaborator issues a new query are the updated saved documents included in the ranking model, and are saved and excluded results hidden. Documents that are promoted by relevance feedback to a previous page are always shown on the current page, to prevent users from missing potentially relevant documents. Note, that in this delayed update model, the status change of the saved documents widget and save/exclude buttons attached to each search result are still immediately occurring.

2.4.7 Software Testing and Logging

Since the SearchX's version V1, we have been coping with software testing, with contributors of SearchX did not have enough knowledge to specify and implement tests. Our initial tests were first implemented in the backend, where the search provider and cache had unit tests. After adding the collaborative features such as synchronisation of recent queries and saved documents, we did not implement unit tests and manually tested if the component was implemented as intended. SearchX's version V4 do not have any unit tests, which could be included using libraries such as the React Testing library [17]. Unit tests can be beneficial to facilitate researchers to implement different components for SearchX.

Another issue with SearchX is reliable logging. We have observed that developers must manually add repetitive logging functions to capture clicks, scroll, and hovers for logging interactions in new components. Besides, logging of mobile devices and tablets are not reliable, and during our user studies, we have thus far disabled the use of SearchX in such devices as we would have to capture touch screen triggers. Therefore, integration with much more reliable logging libraries such as LogUI [89] is still an ongoing project for future development.

2.5 Conclusions

In this chapter, we have presented SearchX, a CSE system whose design and implementation is an ongoing process, born out of the unmet need for an open-source CSE tool that can be deployed online without the need for additional installations (one of the main reasons for ruling out Coagmento for our purposes). SearchX implements essential features of CSE and accommodates extensions for future research. During the development of SearchX we have been faced with many challenges that included failures and success stories when shipping SearchX in CSE user studies using crowd-sourcing platforms and lab user studies.

We have described the versions we developed in this thesis, but there have been a few parallel developments of SearchX that are worth mentioning. Roy et al. [121] incorporated a notepad where users could copy and edit texts from documents and snippets, which allows users to collect, share and make sense of information from their searchers. Roy et al. [121] also modified the document viewer to enable users to highlight pieces of text in the document viewer, which also helps users to share awareness among members. Câmara et al. [28] provided a task scaffolding widget to guide users in their exploratory search to acquire knowledge about a given topic, which can be beneficial to users construct and make sense of the information they find in their search exploration. Salimzadeh et al. [123] developed an entity card widget to measure how much entity cards can support knowledge gain in learning-oriented tasks. Finally, Maxwell and Hauff [89] have provided LogUI library, a library to avoid inconsistent logging issues in web applications such as SearchX, which has been incorporated to SearchX and will provide reliable logging to future researchers.

[17] https://testing-library.com/docs/react-testing-library/intro/

3

On the Impact of Group Size on Collaborative Search Effectiveness

In this chapter, we investigate the first research theme of this thesis on the impact of differences in group configurations in CSE. In particular, we aim to shed light on the effectiveness of CSE when we vary group sizes. To this end, we present a large-scale experiment with more than 300 crowd-workers. We consider the collaboration group size our dependent variable and investigate collaborations between groups of up to six people. We find that most prior (simulation-based) results on the impact of collaboration group size on behaviour and search effectiveness cannot *be reproduced in our user study. The results in this chapter indicate that a further scaling up of group sizes is feasible with existing CSE features and can potentially lead to new research avenues in the CSE space. The contributions of this chapter have been mainly published in [96].*

3.1 Introduction

One dimension that has largely been considered as a constant in CSE studies is the *size* of the collaborating group. Almost all existing CSE studies—no matter if evaluated in a lab or a simulation setting—consider groups of either two [23, 24, 55, 63, 64, 69, 73, 102, 111, 126, 129, 135, 136, 144], three [3, 103] or four [31, 109] users. This strong focus on pairs of collaborators can be explained by the fact that many studies investigate novel CSE features to, for instance, increase awareness of group members' interactions; to facilitate the sharing of knowledge among users; and to employ division of labour. As the number of experimental variants in such studies can increase quickly, a common way to limit the number of variants is to keep the group size constant. In addition, fixing the group size to either two or three users is often motivated by two large-scale surveys that were conducted on the use of CSE in 2006 and 2012 respectively [100, 101]. During that time period, the frequency of CSE episodes increased ten-fold with more than 10% of surveyed users in 2012 reporting daily CSE episodes. When asked about their most recent CSE episode (in the 2012 study), slightly more than half of the participants reported this having been a collaboration between two or three users; at the same time, more than 21% of respondents reported group sizes of five and more. Thus, collaborations between more than three users are not a rare occurrence, though they are hardly investigated in research.

An exception to the observations above is [70] who—in a simulation study—investigated the effect of *changing* group sizes in CSE. Groups of up to five *simulated* users were explored across a range of algorithmic mediation strategies (some of which we also explore in our work); the authors found larger group sizes to lead to higher search effectiveness in a recall-oriented task, albeit with diminishing returns. Simulations though are by definition simplifications of the real world. They ignore the increase in cognitive load real users are likely to experience as group sizes increase and coordination efforts become more difficult to manage. There is thus a lack of evidence of (i) to what extent the simulation findings carry over to a real user study, and, (ii) to what extent the currently existing CSE mechanisms (algorithms as well as interface elements) *scale* to group sizes beyond the commonly investigated sizes of two or three. In order to investigate these issues, we design a crowd-sourcing user study guided by the following overarching research question already introduced in Chapter 1:

RQ1. What is the impact of group size on CSE in terms of retrieval effectiveness and search behaviours?

To answer **RQ1.**, we use our CSE framework `SearchX` (see Chapter 2) with algorithmic mediation (among others, shared relevance feedback) that was found to be effective in prior works. Our crowd-sourcing study with 305 participants based on prior best principles with group size as main dependent variable, and we use groups of 2, 4 and 6 collaborating searchers. Due to the limitation of finding a number of large groups for our study, our end point is set group size of six.

The main contributions and findings of our work are:

- We find most prior simulation-based results on the impact of group size on behaviour and search effectiveness to not hold in our user study with several hundred crowd-workers.

- Importantly, in our study—conducted across three difficult recall-oriented search topics—we do *not* observe diminishing returns (measured in recall) when scaling up group sizes from two to six collaborators. Our results indicate that a further scaling up of group sizes is feasible with existing CSE features and can potentially lead to new research avenues in the CSE space.

3.2 Related Work

In Chapter 2 we have introduced related work on CSE systems, and `SearchX`, our CSE system. In this section, we describe related work that connects to the group sizes in CSE.

As described in Chapter 2, four design principles have been formulated by [99]:

1. *raising awareness* among the collaborators about their activities (e.g. by providing a shared query history);
2. enabling the *division of labour* (e.g. by automatically providing different results to collaborators or enabling collaborators to chat and divide the search task manually);
3. *persistence* (e.g. by storing the query history persistently);
4. enabling *sensemaking* (e.g. by providing multiple views of the common activities).

The first two design principles are most often explored in the information retrieval community. Approaches to raising awareness and dividing the labour can be categorised as belonging to one of three levels [70]: the *interface level* (e.g. a chat widget enables users to divide the work manually), the *techniques level* (i.e. established IR technologies such as document clustering are employed to facilitate the collaboration) or the *ranking model level* (i.e. the ranking model is adapted specifically for the collaborative use case).

We now present prior works in each of these categories in turn and then finish the section with an overview of existing tools and a detailed look at prior works on group size dynamics in CSE. We focus in particular on the ideas presented in prior works, as the reported findings are often based on small user studies—some of those more than ten years old—that explore the use of a single CSE system in a single setting. Table 3.1 presents an overview of key user study statistics (group sizes investigated, number of topics, corpora used, etc.) across a range of user studies; as a comparison, the final row showcases our study.

3.2.1 Interface Level

Morris et al. [103] proposed to raise awareness among collaborators about their querying actions by incorporating visual hints (such as underlying collaborators' query terms) at the search snippet level. In an earlier user study, Morris and Horvitz [102] had shown that a shared query history among collaborators is a significant source of SERP views, with more than 30% of SERPs in the study being retrieved from query history clicks. Instead of just a simple shared query history, the `CoSense` system [109] provides users with a detailed shared timeline (providing information on clicks, views and chat messages) as well as a shared workspace. Capra et al. [31] included filtering options in the SERP, enabling collaborators to view the documents rated as (non-)relevant by their collaborators.

Table 3.1: Summary of key statistics of empirical evaluations of CSE: group size (GS), number of groups (#G), number of search tasks per group (#T) and study type: [sim.] refers to a simulation study with batch evaluation, [lab-fixed] to a lab user study with one or more fixed work/personal search tasks, [lab-nat.] to a lab user study where users self-selected their search task(s). Collection refers to the data collection used. – indicates that it is unknown.

	GS	#G	#T	Type	Collection
Morris and Horvitz [102]	2	7	1	lab-nat.	Web
Amershi and Morris [3]	3	12	3	lab-nat.	Web
Morris et al. [103]	3	10	1	lab-nat.	Web
Pickens et al. [111]	2	4	24	lab-fixed	TRECVid07
Joho et al. [69]	2	12	3	lab-fixed	Aquaint
Paul and Morris [109]	4	12	1	lab-fixed	Web
Joho et al. [70]	1–5	500	13	sim.	Aquaint
Shah et al. [129]	2	5	10	sim.	-
Capra et al. [31]	4	11	1	lab-fixed	Aquaint
González-Ibáñez et al. [55]	2	30	1	lab-fixed	Web
Kelly and Payne [73]	2	8	1-3	lab-nat.	Web
Soulier et al. [136]	2	–	20	sim.	TREC Vol. 4
Soulier et al. [135]	2	70	1	lab-fixed	Web
Tamine and Soulier [144]	2	75	1	lab-fixed	Web
Shah et al. [130]	2–3	34	1	lab-fixed	Web
Htun et al. [63]	2	55	13	sim.	Aquaint
Böhm et al. [23]	2	–	314	sim.	OHSUMED, CLEF-IP
Htun et al. [64]	2	10	3	lab-fixed	Aquaint
Shah et al. [131]	2–3	34	1	lab-fixed	Web
Our work	2–6	67	3	lab-fixed	Aquaint

This, though, did not lead to higher recall levels as often collaborators rated the same documents instead of exploring new areas of the search space. Diriye and Golovchinsky [41] incorporated a search result histogram in the UI of their CSE system, enabling users to keep track of the queries that resulted in a document being retrieved.

Facilitating the division of labour—beyond providing a chat widget as offered in a number of systems, e.g. [102]—has been explored, for example, in the CoSearch system [3] which assumes a shared-computer CSE setting with one main device (e.g. a Desktop) and several smaller devices (e.g. mobile phones) to enable distributed control of the search. The SearchTogether system [102] includes a "recommendation queue" interface feature, enabling users to recommend documents to their collaborators for reading.

3.2.2 Techniques Level

Beyond the interface level, SearchTogether [102] also offers a "split searching" mechanism to distribute the labour (and thus avoid redundancy) with just one of the collaborators submitting a query and the search system splitting the search results in a round-robin fashion across all group members for evaluation. A more intelligent form of splitting was later proposed by Morris et al. [103]: here, each collaborator's personal profile was taken into account in the splitting process. Joho et al. [70] approached split searching by topical clustering with every collaborator receiving the documents associated with a particular aspect of the topic but did not observe increased search effectiveness compared to simple round-robin splitting.

3.2.3 Algorithms Level

While considerable work on the techniques level investigates how best to split up the document space, on the algorithms level we consider changes made to the retrieval algorithms themselves, in particular changes based on asymmetric *user roles*. Here, collaborating searchers are no longer treated as equals but are either assigned fixed [111, 129, 136] or dynamic roles [135] based on their search strategies and behaviours.

Pickens et al. [111] were one of the first works to propose *algorithmic mediation*, formulating two roles in a CSE process, each with their specific ranking algorithm and user interface designed for their respective task: the *prospector* issues diverse queries in order to explore the search space while the *miner* acts as assessor of documents, in particular those occurring highly ranked in many of the prospector's result lists. Two alternative role types that no longer differ in their task (issuing queries vs assessing documents) but in the type of information received were introduced by [129]: the *gatherer* receives result lists optimised for effectiveness while the *surveyor* receives result lists optimised for diversity. In both studies, role-based algorithmic mediation led to a higher search effectiveness than the naive merging of search results by independent pairs of searchers.

Additionally, previous works have shown similar positive results when assigning collaboration roles (i) according to domain expertise [136] or (ii) dynamically based on users' search behaviours [135].

Lastly, [23] have developed a first formal cost model for collaborative result ranking with the aim of deriving (theoretically) optimal collaboration strategies.

3.2.4 Group Size Dynamics

Let us consider once more the work of Joho et al. [70] who investigated the impact of changing group sizes on retrieval effectiveness in a simulation study. Specifically, the query and assessment actions of collaborating groups (one to five users per group) were simulated[1] and eight different search strategies aimed at knowledge sharing and division of labour were evaluated (including independent searching, searching with judged documents removed from the SERP, query expansion with independent/shared relevance feedback, etc.). Increasing group sizes led to increased search effectiveness measured in recall, though with diminishing returns—the largest change in retrieval effectiveness (≈ 50% increase) was observed when a second member entered the team, the smallest when adding the fifth member (5-12% increase depending on the search strategy). However, due to the simulated nature of the study, it is unclear whether those findings will also hold when real users are collaborating—we investigate this very research gap in this chapter.

Additionally, despite not being the focus of this chapter, co-located studies have shown some evidence that increasing group sizes can provide higher group recall in recall-oriented

[1] Note, that the simulation made use of the Robust track 2005 data [155]; we use the same topics and corpus in our experiments.

tasks (groups of three performed significantly better than pairs) [130]. Shah et al. [130] found that groups of three could cover more subspaces of the search space and divide the work more successfully. In this chapter, we also investigate if this holds for groups larger than three.

3.3 Hypotheses

Before describing the study design, we now list the specific hypotheses we investigate in this chapter. We derive our hypotheses based on results found in previous works in [70, 130]—all of them related to the impact of group size on search behaviour, effectiveness and processes.

H1 As group sizes increase, group recall increases, with diminishing gains.

H2 For topics with a higher number of relevant documents, increased group size will significantly impact group recall (as it takes more effort to cover all relevant documents).

H3 Larger groups are more useful at the beginning of the search session, with gains in recall over smaller group sizes decreasing as the search session progresses.

H4 Sharing of knowledge across a group of users increases their group recall; the effect is similar across group sizes.

H5 Larger groups will discover more areas of the information landscape that they might have otherwise missed working by themselves.

H6 The benefits of adding extra members are inversely proportional to the group recall, and it is smaller at the early stages of the search session.

3.4 Study Design

In this section, we present details of our study design, the search variants, SearchX setup, search topics and retrieval models. Then, we describe our experiment in more information, including the crowd-sourced task setup, questionnaires, and evaluation methodology.

3.4.1 Search Variants

To study the impact of different collaboration features we deployed three variants of our search system. The three search system variants we explore are listed in Table 3.2; we also point out their correspondence to the variants discussed by [70]. Variant **S-Single** does not contain any CSE affordance, each searcher receives a single-user search instance. In contrast, variant **S-UI-Coll** provides two interface-based awareness features: a shared recent query and a shared saved-documents widget making all queries and saved documents available to all collaborators (see ④ and ⑤ in Figure 2.3 in Chapter 2). In addition, we also implemented a soft *division of labour* (DoL): results saved by any collaborator are by default hidden from the result listings of all collaborators in the group (though it is possible to "unhide" them); documents explicitly marked as "exclude" by any collaborator are also hidden by default (see Figure 2.7 in Chapter 2). The third variant, **S-UIAlg-Coll** has the same features as **S-UI-Coll** as well as algorithmic mediation for sharing of knowledge. Concretely, we implemented *shared relevance feedback* (RF), where the documents saved by all collaborators are employed in the query expansion stage of all collaborators.

We selected **S-Single** and **S-UI-Coll** in order to determine the benefit of adding more collaborators in a basic CSE setting—instead of designing an interface with as many CSE features as possible, we aimed at a search interface that looks familiar to today's web searchers while still providing awareness and soft division of labour features. We chose not to include a chat widget, as with increasing group sizes the existence of a chat is likely to lead to a long start-up time (users communicating and managing their searches). We chose **S-UIAlg-Coll**, as shared RF has been shown to significantly outperform all other variants in [70]'s simulation study.

In addition, we explore here to what extent the simulation results hold in an experiment with actual users (i.e., crowd-workers). While crowd-workers are only an approximation of "real" users, they are as close as we can get in a large-scale user study. We hypothesise that with users the benefits of shared RF may be outweighed by the cognitive load experienced by users whose search results no longer match their expectations (and this problem is likely to get worse as group sizes increase).

Due to the large number of study participants already required for those three search variants, we opted to not include a fourth search variant with individual relevance feedback.

Table 3.2: Overview of our CSE conditions and their correspondence to the variants explored by [70]. Notice that the judged documents are automatically hidden (though "unhiding" them is possible too).

S-Single	Independent search with individual bookmarks and individual query history (no awareness, no division of labour)
	Similar to *SS1* of [70]'s variant "Team members performs search independently"
S-UI-Coll	**S-Single** + Shared saved documents, shared query history and collapsing of saved and excluded documents in the SERP (awareness, interface-level division of labour)
	Corresponds to *SS2* of [70]'s variant "SS1 with unjudged documents only"
S-UIAlg-Coll	**S-UI-Coll** + Shared relevance feedback (awareness, interface-level division of labour and system-level sharing of knowledge)
	Corresponds to *SS4* of [70]'s variant "SS3 with shared relevance feedback"

3.4.2 SearchX Setup

We used our search system SearchX version V2 (see Figure 2.3 in Chapter 2). For single-user search participants, we provided SearchX with a search box without search verticals ①, a document viewer (see Figure 2.8) without document rating ⑨ and document annotation 10, recent queries group 4, and saved documents 5. For the synchronous CSE of two or more users, we also included the interface elements recent queries group 4 and saved documents 5 but with colour coded for each group member (see Figure 2.6a). Next, we describe implementation details on how we synchronised crowd-workers and algorithmic mediation.

Synchronising Crowd-workers

In lab studies, experimenters often sign up *groups* of collaborating users, e.g. [69, 73, 102, 103, 109, 144], instead of individuals that are grouped together on the fly—those groups are stable, it is unlikely that a member drops out in the middle of the experiment. In a crowdsourcing setup, we can neither sign up groups of workers that know each other nor ensure that every worker completes the task. To overcome these issues we implemented a virtual "waiting room" where crowd-workers who signed up for our task were asked to wait up to 10 minutes (we also offered a game of snake to pass the time). Once the desired number of crowd-workers had signed up, or the ten minutes were up, they received a shared CSE session (randomly assigned to one of the three search variants) and the waiting room became available for the next set of crowd-workers. In the case where the desired number of collaborators was not reached, our system split up the active collaborators into one or more different CSE instances, depending on the number of participating groups required for each group size. Concretely, we initially settled on evaluating collaborations between groups of sizes 1 (i.e. single-user search), 2, 4 and 6 collaborators. If for example after 10 minutes of waiting five workers had joined our virtual waiting room, the system created two groups (one of size 1 and one of size 4).

Lastly, since each group had to tackle three search topics, we could not rely on workers to individually move to the next search topic. We provided workers with a visible timer and the interface switched automatically to the next topic after 10 minutes, ensuring that collaborators remained synchronised in their search topics.

Synchronised Algorithmic Mediation

Shared relevance feedback (utilised in **S-UIAlg-Coll**) and division of labour (utilised in **S-UI-Coll** and **S-UIAlg-Coll**) can be implemented in one of two ways: either immediately or delayed. In the immediate version, as soon as a collaborator saves a document that action should be reflected in the SERPs of all collaborators—not just by updating the shared saved documents list and hiding the document in question from the SERPs, but also by rerunning each user's submitted query with the new set of relevant documents. This is likely to confuse users as they cannot anticipate when (and why) a result ranking suddenly changes; worse still, if a user paginates through the result list, she might miss the newly highly ranked relevant documents because she is looking at lower ranks. We overcome this issue by opting for delayed RF and division of labour: only when a collaborator issues a new query are the updated saved documents included in the ranking model, and are saved and excluded results hidden. Documents that are promoted by RF to a previous page are always shown on the current page, to prevent users from missing potentially relevant documents. Note, that in this delayed update model, the status change of the saved documents widget and save/exclude buttons attached to each search result are still immediately occurring.

3.4.3 Search Topics and Retrieval Models

As corpus we chose TREC AQUAINT (see Chapter 2.3.4). Although it is an older and rather small corpus, it is still a preferred choice for interactive IR studies due to its clean nature (newswire texts). It was also the collection of choice in the TREC 2005 Robust track [155]; and we thus refer to it (TREC AQUAINT plus TREC 2005 Robust track topics) as ROBUST05. It is focused on 50 poorly performing topics in an ad hoc retrieval setting. In order to select the topics for our study, we took the best three automatic runs submitted to ROBUST05 and for each topic computed the mean of the average precision across those runs. We ranked the topics in ascending order and considered only the first ten, i.e. the topics most difficult for the best performing retrieval systems at the time. We chose those topics, as CSE is most appropriate for difficult search topics; note that this choice is in contrast to prior TREC AQUAINT based CSE studies (see Table 3.1) that often opted for *interesting* topics, largely ignoring topic difficulty. As some of those ten topics share relevant documents, we manually selected three very different topics with the additional constraint of at least 30 relevant documents and at most 100 relevant documents in the corpus—too many or too few relevant documents will limit the insights we can gain from our study. We note, that only topic 367 overlaps with the topics employed by [70]. According to our selection criteria, the remaining topics employed in the simulation study were not of sufficient difficulty. Additionally, because we aim to collect evidence for our subspace exploration hypothesis (see hypothesis **H5**), we need subtopics for the topics in ROBUST05. However, subtopics are not provided by the organisers. Thus we considered an extension of the ROBUST05 corpus provided by [90] for topics 341 and 367 and their subtopics. They manually selected subtopics (aspects) for five topics, and two of them (341 and 367) overlapped with the topics we chose for this study. Finally, as described in Chapter 2.3.4, we removed duplicated and invalid documents from TREC AQUAINT, which also included relevant documents in the original ROBUST05. The final three topics are listed in Table 3.3, with the number of relevant documents before and after cleaning and the number of subtopics per topic acquired from [90].

Table 3.3: ROBUST05 topics selected for our study, including the number of relevant documents in TREC AQUAINT (after and before cleaning duplicates and invalid documents) and the average AP of the three top performing ROBUST05 runs. Last column refers to the number of subtopics (aspects) provided by [90].

ID	**Topic**	**#Rel. original**	**#Rel. after cleaning**	**av. AP**	**# Subtopics**
341	airport security *Example subtopics:* *lagos airport* *los angeles international airport* *miami international airport*	37	32	0.08	14
367	piracy *Example subtopics:* *gts katie, virgin pearl* *marine master*	95	81	0.09	17
650	tax evasion indicted	32	24	0.09	-

In variants **S-Single** and **S-UI-Coll**, the retrieval algorithm is language modelling (LM) with Dirichlet smoothing [168] with hyper-parameter setting $\mu = 2500$. The relevance feedback (RF) variant **S-UIAlg-Coll** implements relevance-based language modelling [83], in particular RM2 with 10 feedback terms and all documents saved by the collaborators. In an offline experiment we found RM2 with *true* RF (as found in the official TREC relevance judgement file) to outperform the LM baseline on average by 82.65% across our three topics: we sampled 5 relevant documents per topic from the official qrels 20 times, retrieved ranked lists of results based on RM2 and LM, removed those 5 relevant documents from the result lists (and the relevance judgement file) and computed the recall. Thus, as long as our participants save (mostly) relevant documents, RF will improve the quality of the search results compared to the non-RF based language modelling variant employed in **S-Single** and **S-UI-Coll**.

3.4.4 Crowd-sourced Task Setup

We opted for a crowdsourcing setup due to the number of participants we require: three search variants, each evaluated with ideally four group sizes (1/2/4/6). We aim for ten groups in each setup—a common size, cf. Table 3.1—thus requiring more than 300 participants, considerably more than in any of the listed lab studies.

We recruited workers from the Prolific platform[2], which has been shown to be a more reliable source of workers for cognitively demanding tasks than MTurk or FigureEight (formerly CrowdFlower), two other popular crowdsourcing platforms [110]. Workers can only participate once in our study. Each worker, once accepting the task, is directed to our server and moves through the workflow depicted in Figure 3.1. The pre-task questionnaire contains a description of CSE and four CSE questions borrowed from [101] to prime the workers for the upcoming CSE tasks:

1. How often do you engage in collaborative Web search? *5-Likert scale [Daily, Less often]*

2. Describe what were you looking for. (e.g. husband and wife planning a trip for the family, a group of students working on a writing assignment and sharing search results/findings, a couple shopping for a new sofa, etc.)

3. With how many others did you collaborate (not including yourself)? *[Number]*

4. It was easy to determine if a document was relevant to a task. [Disagree, Agree]

The waiting room component contains an explanation of the "waiting room" concept, a visible timer, the option to play a game of snake and the option of an audio alert, to enable the worker to use other browser tabs while waiting for sufficiently many workers to join. Once workers move to the CSE phase, they first receive an interactive tour of the search interface before starting to work on their assigned topics, which we randomised the order that they were provided to mitigate learning effects in our analyses. To provide context for the ad-hoc search topics, we employed the task template in Figure 3.2, inspired by previous studies [11, 80]. After 10 minutes of searching, all collaborators are automatically moved to the next search task to ensure synchronisation. This recall-oriented task can be found in

[2] https://www.prolific.co/

settings such patent-retrieval and e-discovery and represents a typical task in CSE which induces a complex and exploratory behaviour from users as discussed by [101].

The post-task questionnaire contains the following seven questions on search satisfaction typically asked in interactive IR studies [72]. We also asked questions to evaluate the collaborative features we provided from `SearchX`:

1. How many people did you just now collaborate with (not including yourself)? *[Number]*
2. The color coding of the query history and bookmarks made sense to me. *5-Likert scale [Disagree, Agree]*
3. It was easy to understand why documents were retrieved in response to my queries. *5-Likert scale [Disagree, Agree]*
4. I didn't notice any inconsistencies when I used the system. *5-Likert scale [Disagree, Agree]*
5. It was easy to determine if a document was relevant to a task. [Disagree, Agree]
6. How difficult was this task? *5-Likert scale [Very easy, Very difficult]*
7. Did you find the collaborative features useful (multi-grid question with one row for each feature: recent queries, saved documents, and hiding saved and excluded results)? *5-Likert scale [Disagree, Agree]*

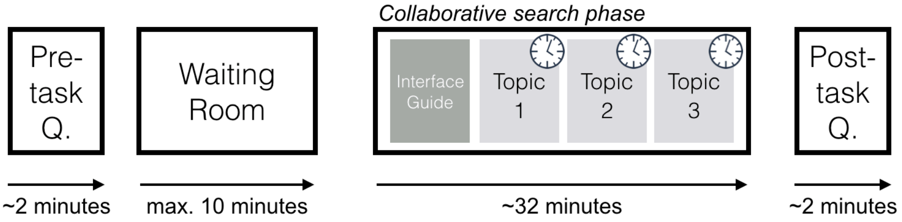


Figure 3.1: Overview of a worker's flow through our system.

> Imagine you are a reporter for a newspaper. Your editor has just told you to write a story about [`ROBUST05` **topic title**]. There's a meeting in an hour, so your editor asks you and your colleagues to spend 10 minutes together and search for as many useful documents (news articles) as possible and save them. Collect documents according to the following criteria: [`ROBUST05` **topic description**].

Figure 3.2: Task template for our study.

Workers are assigned to search variants and group sizes at random; the order of the three search topics is randomised per group.

3.4.5 Post-processing of Collected Logs

Due to the unpredictable nature of crowd-workers, a synchronised CSE experiment is not easy to conduct. As group sizes increase, it becomes more difficult to form groups (as sufficiently many crowd-workers have to choose the task at roughly the same time) and worker dropout becomes more likely during the task. We mitigate these issues in three steps:

1. On a *per topic* basis, we only consider collaborators as *active* in a group that issued at least one query for the topic.

2. We consider the number of collaborators in a group on a topic-by-topic basis (we only count active collaborators), instead of fixed across all three search topics. For instance, if a group starts with four workers, and one worker becomes inactive after the first search topic and another worker drops at the start of the third topic, we consider this as a group of four collaborators for the first topic, a group of three for the second topic and a group of two for the third topic.

3. As after these two steps we have groups of three and five collaborators, we subsume the logs of groups of 3-4 and 5-6 collaborators respectively into two groups in the analyses that follow. Overall, we thus analyse four group sizes: $\{1, 2, \{3, 4\}, \{5, 6\}\}$. We decided on this merging strategy (instead of for instance merging together groups of 2-3 participants) as [70] found in their simulation study the addition of the second collaborator to bring about the greatest benefit in terms of search effectiveness.

Table 3.4 provides an overview of how many groups in total (67 groups with between 2 and 6 participants—recall that a group has three topics to participate in) participated in our experiment across all search variants after the above post-processing steps. As **S-Single** is the single-user search setup, we collect 12 instances of single-search tasks and then *simulate* the behaviour across larger group sizes by grouping users together and merging their saved documents in line with prior works [47, 63, 70, 111]. We group users together by considering all possible combinations for each group size, ensuring that the data for each user is weighed equally in the results.

A total of 335 workers participated in our study, of which 30 were excluded since they did not perform any actions. We thus had 305 valid participants. In Table 3.4 we have the largest number of groups (16) for condition **S-UIAlg-Coll** and groups $\{3, 4\}$. This artefact can be explained by the fact that regularly participants assigned to collaborating in a team of six (which required the longest waiting time) grew impatient and dropped out, and thus often the result were formed groups of three or four participants. On average, our workers spent 42 minutes on the task, including the at most 10 minutes in the virtual waiting room. We paid £3.75 for the task, which was above the price suggested by the Prolific platform.

The drop-out rate of participants *during* the experiment was 30.4%, that is the rate of groups starting off with at least two collaborators that decreased in size whilst working through the three topics.

3.4.6 Evaluation Metrics and Statistical Tests

We use the following metrics and statistical analyses to compare the search variants and group sizes in line with previous works by Joho et al. [69] and Pickens et al. [111].

Table 3.4: Number of groups across search variants, topics and group sizes. Number of groups across search variants, topics and group sizes. For **S-Single**, we created artificial groups by simulating the CSE behaviour across larger group sizes with the data collected from the single-user search data.

	Topic ID	{1}	{2}	{3,4}	{5,6}
S-Single	650	12	–	–	–
	367	12	–	–	–
	341	12	–	–	–
S-UI-Coll	650	11	12	10	9
	367	12	11	10	9
	341	13	10	11	8
S-UIAlg-Coll	650	17	8	16	12
	367	17	11	13	13
	341	19	10	14	13

Retrieval effectiveness: in order to measure retrieval effectiveness, we employ group recall which is defined as the recall for the union of the sets of documents that each collaborator saved in comparison with documents provided relevant by TREC 2005 Robust track [155]. This metric is appropriate, given the fact that we designed our search tasks to be recall-oriented. The group recall $GR(g,t)$ is calculated for each group g and topic t. To calculate the average group recall $AGR(s,t)$ for all groups for size s and topic t, we average group recall across all groups in a given size category:

$$AGR(s,t) = \frac{1}{|G_{s,t}|} \sum_{g \in G_{s,t}} GR(g,t). \tag{3.1}$$

where $GR(g,t)$ is the group recall measured as the number of relevant documents retrieved by a group g for topic t divided by the number of relevant documents in the collection for topic t.

We also considered average group subtopic recall ($AGSR$). To calculate $AGSR(s,t,S_t)$ for all groups for size s and topic t and its subtopics S_t, we average group subtopic recall all groups in a given size category:

$$AGSR(s,t,S_t) = \frac{1}{|G_{s,t}|} \sum_{g \in G_{s,t}} GSR(g,t,S_t). \tag{3.2}$$

where $GSR(g,t,S_t)$ is the number of subtopics retrieved by group g for topic t divided by $|S_t|$, which is number of subtopics for topic t.

Temporal analyses: at a number of time points during a search session—specifically, after each minute—the group recall for each group is computed; here, we only take documents saved until that point in time into account. The search session start time is fixed to the time the first member of a collaborating group submits a query. Average group recall for different group sizes is computed in the same manner as discussed above.

Statistical analyses: in order to compare the impact of group size and search variant in Section 3.5, we conducted a two-factor analysis of variance (ANOVA) separately by each topic. We examined the ANOVA assumptions with Levene's test (homogeneity of variances) and Shapiro-Wilk's test (normality of the ANOVA residuals). We conducted a *post-hoc* analysis using Kruskal-Wallis for the **S-Single** search variant and Tukey's HSD test for the other search variants.

3.5 Results

We now first present an overview of the main outcomes of our study and provide insights into our participants' search behaviours. We then discuss these results in light of the research hypotheses listed in Section 3.3.

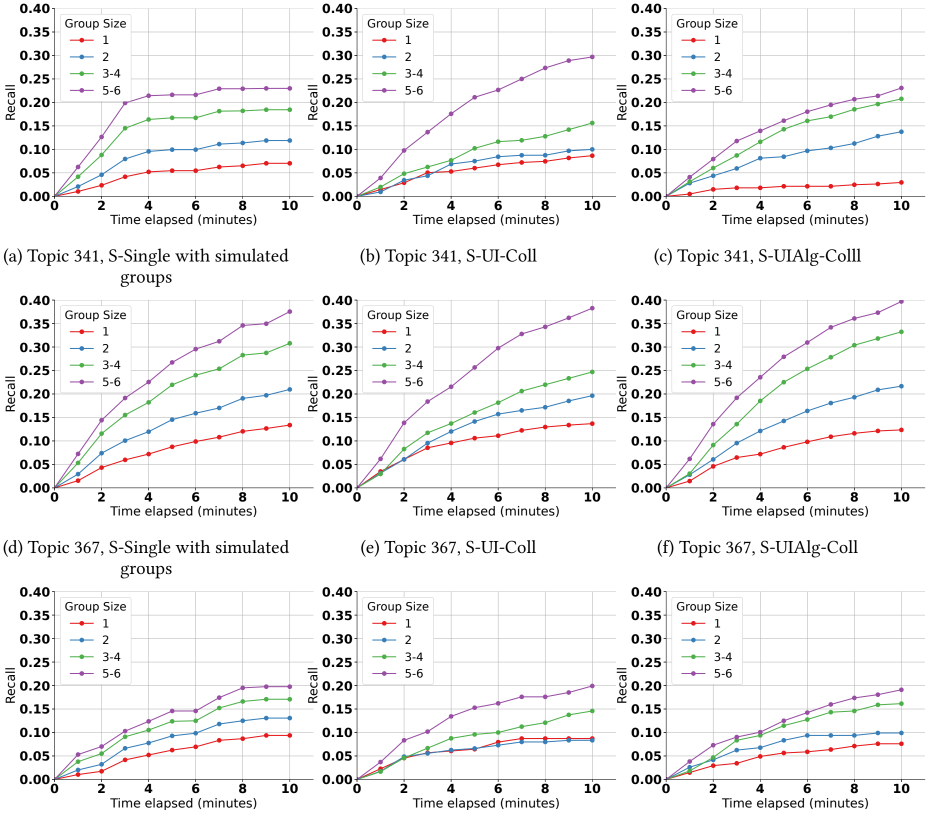
The main results of our study are shown in Figure 3.3 where we present a detailed overview of the development of average group recall across time, different group sizes and search variants. Two main observations can be made:

- in all cases a larger group size leads to a larger recall level;
- in line with Joho et al. [69]—and in contrast to [126]—we did not observe a synergy effect: pairs of collaborators were *not* more effective than two independent searchers whose results were aggregated.

Table 3.5 lists the average group recall on a per-topic basis, with statistical differences in recall across group sizes and search variants highlighted—note that due to the nature of **S-Single** (where group sizes > 1 are simulated) it is not possible to reliably test for statistical differences between **S-Single** and **S-UI-Coll**/**S-UIAlg-Coll**. We find that group sizes of $\{3, 4\}$ and $\{5, 6\}$ respectively to lead to significantly higher recall levels than smaller groups. Next, we present our results and provide evidence supporting or not the hypotheses that aim to answer our research question.

3.5.1 Search Behaviours

To provide insights into participants' search behaviours as well as to ascertain that our participants conducted searches as intended, we list major characteristics of their individual behaviour in Table 3.6. In order to aggregate participants' behavioural traces in a meaningful manner, we resorted to computing the median value across all participants of a certain collaboration group size, topic and search variant. Since we have three topics in total, we then computed the average of the three median values. We observe across search variants and group sizes, that the median number of queries issued by a searcher for a topic varies between 5 and 9 with queries being of moderate length (3 to 4 terms). We find a decreasing trend in the number of viewed documents as well as the amount of time spent on each document viewed for participants in larger collaborating groups—there are two possible explanations for this trend: on the one hand, the participants may be more occupied with the activities of their collaborators (as they appear in the shared query history and the shared bookmarking widgets) or, on the other hand, the participants may be more complacent, knowing that their collaborators are active in the same search task. The last column in Table 3.6 shows that complacency is not a likely explanation, as the number of unique saved documents remains relatively stable across collaborative group sizes (with the exception of the single-user case). Finally, we note that the document viewing time is rather short (between 8 and 13 seconds in Table 3.6); this can be explained by the



(a) Topic 341, S-Single with simulated groups

(b) Topic 341, S-UI-Coll

(c) Topic 341, S-UIAlg-Colll

(d) Topic 367, S-Single with simulated groups

(e) Topic 367, S-UI-Coll

(f) Topic 367, S-UIAlg-Coll

(g) Topic 650, S-Single with simulated groups

(h) Topic 650, S-UI-Coll

(i) Topic 650, S-UIAlg-Coll

Figure 3.3: Overview of the average group recall for each topic and search variant measured in one-minute time intervals.

fact that the TREC AQUAINT documents themselves are relatively short, with an average length of 438 words.

Based on these statistics and a manual check of a sample of search logs our participants generated we conclude that our participants provide valid log traces for our analyses.

3.5.2 Search Effectiveness across Group Sizes

From Table3.5, we make our first observation with respect to recall: for none of the topics, search variants and group size combination is the reported recall greater than 0.4, indicating the difficulty of the topics and the potential benefit an increasing collaborator pool could bring about. We also see that, despite picking the most difficult topics, the maximum recall varies considerably (Figure 3.3), with a maximum of 0.2 (topic 650), 0.3 (topic 341) and 0.4 (topic 367) respectively.

While the trends across topics and search variants in Figure 3.3 are similar, there are two apparent outliers: Figure 3.3b shows a significant recall gap (the recall doubles) be-

Table 3.5: Topic-wise average group recall (averaged across all groups in a single topic/search-variant) across conditions and group sizes. Statistical significance was determined via Tukey's HSD test independently for each topic; in each topic column, significant improvements at $p < 0.01$ with Bonferroni correction are marked with superscript ^{XY}. X is the search variant ('I' in the case of **S-UI-Coll** and 'A' in the case of **S-UIAlg-Coll**), and Y is the respective group size. For the **S-Single** simulated groups, we determined significant values among group sizes only within **S-Single** via Kruskal-Wallis test independently for each topic (we omitted superscript symbols as all group sizes show significantly different results at $p < 0.01$).

	Group size	**Average group recall per topic**		
		650	**367**	**341**
	1	0.094	0.134	0.070
S-Single	2	0.131	0.210	0.119
	3-4	0.171	0.308	0.184
	5-6	0.198	0.376	0.230
	1	0.087	0.137	0.087
S-UI-Coll	2	0.083	0.196	0.103
	3-4	0.146	0.249	0.159^{A1}
	5-6	$0.208^{I1,I2,A1}$	$0.391^{I1,I2,A1}$	$0.305^{I1,I2,I34,A1,A2}$
	1	0.076	0.125	0.031
S-UIAlg-Coll	2	0.109	0.231	0.138
	3-4	$0.169^{I2,A1}$	$0.349^{I1,I2,A1}$	$0.214^{I1,A1}$
	5-6	$0.219^{I1,I2,A1,A2}$	$0.404^{I1,I2,I34,A1,A2}$	$0.243^{I1,I2,A1}$

tween group sizes of $\{3,4\}$ and $\{5,6\}$ and Figure 3.3h shows no change in recall between group sizes of one and two. We did not observe anomalies in the search logs across those two topic/variant setups; based on this finding we argue that these are slight variations are a result of our user study setup.

Across all search variants and topics we find the first part of **H1** (*As group sizes increase, group recall increases ...*) to be supported. In **S-Single** each additional group member results in a similar absolute increase in recall levels at the end of the search session, i.e. at the ten minute mark in Figure 3.3.

In contrast, for the interface-based collaboration variant **S-UI-Coll** we find only small differences in recall level between teams of one (i.e. a single searcher) and teams of two collaborators at the end of the search session. The largest increase in recall (ranging from +42% to +92% depending on the topic) occurs when moving from groups of $\{3,4\}$ to groups of $\{5,6\}$ collaborators. The significance tests reported in Table 3.5 show that within **S-UI-Coll** for all topics the largest group size yields significantly better recall levels than group sizes of 1 and 2. In case of topic 341 the difference is also significant with respect to group size $\{3,4\}$ for the **S-UI-Coll** variant.

A somewhat different picture once more emerges when considering algorithmic mediation (**S-UIAlg-Coll**): here, we find smaller differences in recall (between +14% and +30% depending on the topic) when moving from $\{3,4\}$ collaborators to $\{5,6\}$ compared to moving from two to $\{3,4\}$ collaborators (+51% to +55%). These findings show that the

Table 3.6: Overview of individual search behaviours across the search variants. GS is the group size and # refers to number (e.g., # Queries), and Rel. Docs. refers to relevant documents for the topics. For each topic, the mean value is computed; reported here is the average and standard deviation of those three mean figures.

Variant	GS	# Queries	# Unique Queries	Avg. Query Length	# Unique Query Terms	Total Dwell Time	Avg. Dwell Time	# Viewed Docs.	# Unique Viewed Docs.	# Unique Rel. Viewed Docs.	# Saved Docs.	# Rel. Saved Docs.	# Unique Snippets	# Unique Rel. Snippets
S-Single	1	6.47 (3.76)	6.31 (3.63)	4.47 (1.77)	11.22 (6.79)	147.25 (117.70)	16.67 (13.97)	10.31 (5.59)	9.61 (4.98)	3.25 (2.13)	11.78 (4.60)	5.11 (2.13)	9.44 (2.93)	9.44 (2.93)
S-UI-Coll	1	8.29 (5.18)	7.81 (4.98)	4.47 (3.08)	12.12 (6.67)	70.55 (80.28)	8.53 (8.24)	4.90 (4.44)	4.71 (4.20)	1.33 (1.31)	12.74 (9.18)	4.95 (3.25)	9.07 (4.00)	9.07 (4.00)
	2	9.19 (4.29)	8.76 (4.39)	4.26 (1.51)	16.44 (7.78)	52.15 (51.27)	8.25 (6.48)	4.70 (4.57)	4.49 (4.24)	1.11 (1.23)	8.84 (5.88)	3.98 (3.46)	11.63 (7.59)	11.63 (7.59)
	3-4	8.98 (4.83)	7.84 (4.06)	4.16 (1.77)	17.18 (10.36)	70.46 (72.76)	9.36 (9.25)	6.11 (5.50)	5.90 (5.23)	1.31 (1.79)	8.55 (7.89)	2.61 (2.57)	10.15 (6.77)	10.15 (6.77)
	5-6	10.13 (4.93)	9.51 (4.65)	4.81 (1.79)	20.72 (9.75)	58.08 (58.22)	7.58 (6.63)	5.63 (6.54)	5.29 (5.85)	1.09 (1.29)	8.40 (6.56)	2.79 (2.38)	11.97 (5.35)	11.97 (5.35)
S-UIAlg-Coll	1	6.75 (3.20)	6.31 (2.96)	4.13 (1.93)	12.73 (6.76)	135.85 (124.85)	10.10 (7.37)	10.87 (8.29)	10.06 (7.36)	2.98 (2.82)	7.95 (4.28)	3.88 (2.25)	9.84 (4.48)	9.84 (4.48)
	2	8.36 (8.03)	6.90 (3.82)	3.28 (0.49)	14.08 (8.36)	101.59 (83.89)	9.11 (7.48)	10.50 (9.65)	9.87 (8.88)	3.94 (5.11)	9.29 (5.76)	4.85 (4.01)	15.79 (7.95)	15.79 (7.95)
	3-4	7.24 (3.36)	6.83 (2.96)	4.88 (2.05)	17.07 (8.78)	60.21 (71.28)	7.06 (5.52)	5.86 (6.39)	5.54 (6.13)	1.35 (1.57)	8.96 (8.55)	3.35 (2.50)	12.22 (5.18)	12.22 (5.18)
	5-6	9.71 (6.40)	8.70 (4.78)	4.42 (1.63)	22.01 (10.00)	64.85 (75.78)	8.49 (11.88)	5.73 (5.20)	5.41 (4.89)	1.51 (1.57)	8.48 (8.92)	2.84 (2.29)	15.47 (4.93)	15.47 (4.93)

second part of **H1** (... *with diminishing gains*) does not seem to hold when moving from a pure simulation study to a user experiment (i.e., not all three search variants, topics, and groups sizes). In line with our findings for **S-UI-Coll** we here also observe statistically significant differences in recall between the largest group size and group sizes 1 and 2. We also note that in contrast to Joho et al. [70] we do not observe a convergence of the recall levels across group sizes towards the end of the search session—that is to say, while we evaluated no more than six collaborators (as crowd-sourcing becomes more difficult with increasingly high group sizes), based on our results we can expect even larger group sizes to yield additional increases in recall. One explanation for this difference can be found in the fact that we focused specifically on difficult topics, while in Joho et al. [70]'s simulation study easy topics are included which reach high recall levels even across a small number of collaborators.

We now move on to hypothesis **H2**, which states: *For topics with a higher number of relevant documents, increased group size will significantly impact group recall.* Topic 367 has 81 relevant documents, more than double that of topics 341 and 650 with 32 and 24 relevant documents respectively (cf. also Table 3.3). When we consider the recall developments in Figure 3.3 across the different topics, we seem to find that **H2** may not hold: for topic 367 we do not observe a higher impact on group recall than for the other two topics. Our results indicate that the choice of collaboration (interface-based only vs. algorithmic) is a more important factor with respect to explaining the impact of group size on group recall level changes.

3.5.3 Search Effectiveness Across Time

Hypothesis **H3** is concerned with the development of recall over time, with simulation studies indicating that a large group size is beneficial in particular early on in the search process. We restate it here: *Larger groups are more useful at the beginning of the search session, with gains in recall over smaller group sizes decreasing as the search session progresses.* Once again we consider the recall developments in Figure 3.3. We find that in practice the benefit remains relatively consistent, i.e., it is not restricted to the early minutes of the search session. As time progresses, the smaller collaborating groups do not “catch up” with the larger groups in terms of recall; this behaviour holds across all three topics and search variants. Although we do not know what would happen after the ten minute mark (as we fixed the end of a topic's search session), we observe the recall curves for group sizes of 1 and 2 to level off somewhat across most topics and search variant combinations, in contrast to the larger group sizes.

3.5.4 Sharing of Knowledge

Let us now consider hypothesis **H4** which states: *Sharing of knowledge across a group of users increases their group recall; the effect is similar across group sizes.* Recall that **S-UI-Coll** provides interface-level division of labour while **S-UIAlg-Coll** provides both interface-level division of labour and algorithmic sharing of knowledge via shared RF. In contrast to our expectations (simulations found sharing of knowledge to increase group recall compared to post-hoc merging of result lists as done in **S-Single**) we find **S-Single** to perform on par with **S-UI-Coll** and **S-UIAlg-Coll** in terms of retrieval effectiveness. While providing sharing of knowledge (**S-UI-Coll** vs. **S-UIAlg-Coll**) does not yield significant changes in recall level for a given topic and group size, we find that for 8 out of 9 topic/group-size comparisons (ignoring groups of size one) **S-UIAlg-Coll** reports a higher recall level than **S-UI-Coll**, providing some support for **H4**.

Lastly, we determine whether our participants actually engaged with their collaborators through our interface affordances. To this end, in Table 3.7 we report the median number of click interactions our groups of collaborators had with queries from the `Recent Queries` and documents from the `Saved Documents` widgets. Here, we ignore interactions of collaborators with their own posed queries and their own saved documents. We find that with increasing collaboration group size more such interactions take place, though overall their number remains small (e.g. for groups of size {5,6} the median number of query widget interactions is 4 to 5, depending on the topic). This is in contrast to the findings by Morris and Horvitz [102] who reported in the evaluation of their `SearchTogether` system that more than 30% of SERP views could be traced back to query history clicks. It should be noted though, that `SearchTogether` was a standalone Desktop client with very elaborate collaboration widgets, while we strove to make the search experience collaborative but still very much relatable to modern web search.

Table 3.7: Usage of collaborative search interface features by groups of collaborators. We measured only clicks on recent queries and saved documents widgets by collaborators that did not issue (save) the original query (document). For each topic, the median value is computed; reported here is the average of those three median values.

	Group size	#Clicked queries	#Viewed saved docs.
S-UI-Coll	2	0.00	0.00
	3-4	0.50	0.67
	5-6	4.00	1.50
S-UIAlg-Coll	2	0.00	0.00
	3-4	0.33	0.00
	5-6	4.33	2.00

3.5.5 Effectiveness of Exploring the Search Subspaces

Next, we look at hypothesis **H5**, concerned with how groups explore the search subspaces of a search topic. These subspaces can be seen as the subtopics for a search topic. For instance, for a search topic *airport security*, subtopics can be airport security in Los Angeles international airport. We restate **H5** here: *Larger groups will discover more areas of the information landscape that they might have otherwise missed working on by themselves.* As described in Section 3.4.3, we obtained the subtopics for two topics, 341 and 367, and we considered these subtopics the possible *areas of information landsapce* that a group may explore during the search process for a topic. Thus, to measure the proportion of subtopics retrieved during the task, we measured average group subtopic recall (see Section 3.4.6), which is measured by the number of relevant subtopics found in saved documents. Maxwell et al. [90] annotated the relevant documents and generated a list of subtopics relevant for each document, where each document can be relevant for one or more subtopics. Table 3.3 we show examples of subtopics for topics 341 and 367. To simplify our analysis, we consider that each subtopic has the same level of difficulty. In Table 3.8, we show the average group subtopic recall for topics 341 and 367.

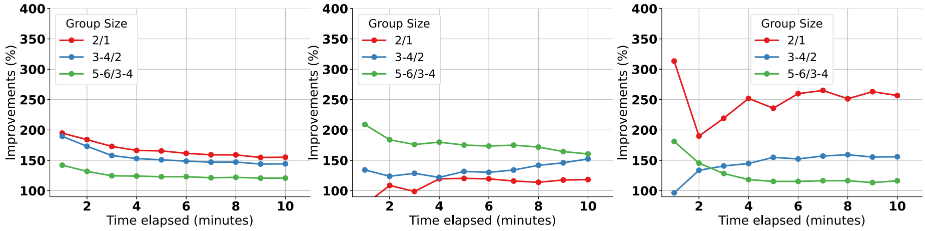
From this table, the first observation we make is with respect to the numbers of subtopics discovered. Collaborating groups discovered more subtopics than individual users, and larger groups (group sizes of 3,4 and 5,6) are significantly more effective in finding new subtopics when they save documents that contained those subtopics. Additionally, similarly to group recall, as we increase group sizes, there is a trend showing that group subtopic recall increases. However, the sharing of knowledge provided in variant **S-UIAlg-Coll** does not seem to help in topic 341. A possible explanation is that topic 341 has fewer relevant documents than topic 367 (32 and 81 documents, respectively), and thus shared relevance feedback made groups discover fewer subtopics. In contrast, for the case of topic 368, finding subtopics seemed to be more effective in terms of group subtopic recall when the number of relevant documents is large. These results show that there is preliminary evidence to support **H5**.

Table 3.8: Average subtopic group recall (averaged across all groups in a single topic/search-variant) across conditions and group sizes for topics 341 and 367. Statistical significance was determined via Tukey's HSD test independently for each topic; in each topic column, significant improvements at $p < 0.05$ with Bonferroni correction are marked with superscript ^{XY}. X is the search variant ('I' in the case of **S-UI-Coll** and 'A' in the case of **S-UIAlg-Coll**), and Y is the respective group size. For the **S-Single** simulated groups, we determined significant values among group sizes only within **S-Single** via Kruskal-Wallis test independently for each topic (we omitted superscript symbols as all group sizes show significantly different results at $p < 0.01$).

Topic	**Group size**	**Average subtopic group recall per topic**		
		S-Single	**S-UI-Coll**	**S-UIAlg-Coll**
341	1	0.167	0.148	0.086
	2	0.239	0.214	0.193
	3-4	0.291	0.292	0.260 $^{\text{A1}}$
	5-6	0.315	0.375 $^{\text{I1,I2, A1, A2}}$	0.264 $^{\text{I1,I2, A1}}$
367	1	0.338	0.309	0.343
	2	0.450	0.433	0.497
	3-4	0.554	0.459	0.620 $^{\text{A1}}$
	5-6	0.620	0.614 $^{\text{I1, A1}}$	0.719 $^{\text{I1, A1}}$

3.5.6 Benefits of Adding New Members

Our last hypothesis **H6** is about the benefits of adding new members to a group, which states: *The benefits of adding extra members are inversely proportional to the group recall, and it is smaller at the early stages of the search session.* This hypothesis comes from the results found by Joho et al. [70]. To show evidence for this hypothesis, we show in Table 3.9, the improvements of group recall by increasing group size. Based on the results reported in Table 3.9, we cannot support the first part of **H6**: *The benefits of adding extra members are inversely proportional to the group recall.* Hence improvements in group recall are not inversely proportional to group size for all topics in the system variants **S-UI-Coll** and **S-UIAlg-Coll**. However, for the simulated groups, improvements in recall seemed to be inversely proportional to group sizes, which is on par with Joho et al. [70]'s simulation



(a) S-Single with simulated groups

(b) S-UI-Coll

(c) S-UIAlg-Colll

Figure 3.4: Overview of the improvements in average group recall for search variant averaged over topics and measured in one-minute time intervals.

Table 3.9: Topic-wise benefit of new members across of group size in terms of group recall (averaged across all groups in a single topic/search-variant) across conditions and group sizes. Note that 3-4/2 means group recall of groups of size 3,4 divided by group recall of groups of size 2.

		Improvements (%)		
Topic	**Group Size Ratio**	**S-Single**	**S-UI-Coll**	**S-UIAlg-Coll**
	2/1	139.39	95.65	143.95
650	3-4/2	130.69	175.00	154.76
	5-6/3-4	115.70	142.86	129.23
	2/1	156.78	143.54	185.10
367	3-4/2	146.85	126.97	151.16
	5-6/3-4	121.95	156.77	115.49
	2/1	169.02	119.17	440.00
341	3-4/2	155.22	154.27	155.84
	5-6/3-4	124.61	191.52	113.30

work. In Figure 3.4, we plot the improvements in recall along with search session the average session across all topics. Here, we find that simulated groups do not to support the rest of **H6** *and it is smaller at the early stages of the search session.* Once again, it is not the case for our real groups in system variants **S-UI-Coll** and **S-UIAlg-Coll**.

3.5.7 Worker Perceptions

Finally, we analyse the participants' responses to the seven questions in the post-questionnaire (the questions are listed in Section 3.4.4). For our analysis, we only consider responses from participants in stable groups, i.e. groups that maintained their original size across all three search topics (as one of the questions is concerned with the *perceived* group size). We thus report responses from 12, 31, and 43 participants in search variants **S-Single**, **S-UI-Coll** and **S-UIAlg-Coll**, respectively. For the post-questionnaire questions 2-7, we observed similar results in terms of median figures, and could not find statistical difference across group sizes and search variants. The most interesting finding pertains to question

1, the perceived vs. actual group size[3] as reported in Table 3.10: while for small group sizes (single-user search or pairs of collaborators) almost always the majority of participants is able to pick the correct group sizes, with increasing group size the perception varies widely, with the vast majority of participants underestimating the size of the group at actual group size {5-6}. Depending on the search variant, we also see a considerable number of participants in the single-user search condition to report themselves having been in a collaboration—we attribute this to the priming questions on CSE, the participants had received at the start of the experiment as well as the virtual waiting room time the participants experienced.

Table 3.10: Group size vs perceived group size in % across search variants and group sizes. Results listed here are based on the post-questionnaire (question 1, Chapter 3.4.4). Shown in underline is the cell value where actual=perceived group size.

Condition	**Group Size**	**Perceived Group Size in %**						
		1	2	3	4	5	6	7+
S-Single	1	50	42	0	8	0	0	0
S-UI-Coll	1	90	0	0	10	0	0	0
	2	6	81	13	0	0	0	0
	3-4	4	23	31	15	19	4	4
	5-6	3	11	20	43	9	6	8
S-UIAlg-Coll	1	69	19	0	0	6	6	0
	2	0	44	19	25	6	0	6
	3-4	6	26	35	12	15	6	0
	5-6	4	19	10	45	12	10	0

Questions two to six in our post-questionnaire focus on the participants' search experience, specifically (Q2) color coding, (Q3) easy to understand document retrieval, (Q4) no inconsistencies, (Q5) easy to determine relevance, (Q6) task difficulty. The feedback for questions Q2 to Q5 is similar across the search variants and group sizes, ranging from 3-4 on the 5-point Likert scale, and thus we can conclude that the general search experience was positive. For Q6, we also find similar task difficulty values across the three search variants ranging from average values of 2.75 (**S-Single**) to 2.79 (**S-UIAlg-Coll**) and 2.84 (**S-UI-Coll**).

With respect to the usefulness of the different interface features (Q7), for all the search variants, participants agreed the saved documents to be the most useful feature, followed by the recent queries and then the hiding of already saved/excluded search results.

3.6 Limitations

In this section, we analyse the limitations of this chapter as our experimental setup had limitations related to: (i) the remote CSE setting; (ii) paid participants to conduct not fully

[3] Note that while the question in the questionnaire asked the participant for the number of collaborators *not including him/herself,* we here report the perceived group size number to simplify the comparison.

realistic tasks; (iii) user study instructions; and (iv) the number of task and the nature of the tasks we gave our participants. First, we acknowledge that results for larger group sizes may not generalise to the co-located setting across all CSE dimensions. In particular, we limited the communication dimension by disabling the chat tool, which—in the co-located setting—is hard to control. Also, for the **S-Single**, we did not provide the same interface (e.g., collapsing of saved and excluded documents), which is was provided in the other two variants. We believe that this is a limitation to our work, and results could have been impacted, such as the behaviour of the **S-Single** to stop their searches and doing more pagination actions. Second, we acknowledge that participants are paid to take part in a study and have not themselves encountered the tasks we gave them, but we claim that it does not affect the importance of the presented results but leaves space for future work. Third, our user study is designed following instructions based on previous work in collaborative search and guidelines in HCI for information retrieval [72, 128, 144]. We acknowledge that the instructions were given to our participants as a limitation to this study, and future work should investigate the impact of such instructions, which is out of the scope of this thesis. Lastly, we designed our tasks aimed at being as realistic as possible, however, we acknowledge that they are still controlled and somewhat artificial tasks. Also, we only selected three topics and hard topics for our study. These topics and datasets are not representative of an up-to-date data collection that participants would find in their daily searches on the Web. Besides that, our tasks are recall-oriented tasks, which many participants may not have encountered before in a general collaborative search setting. We acknowledge that, in combination, all these limitations could threaten the generalisability of any findings, but we believe that this does not devalue the results found in this chapter.

3.7 Conclusions

The impact of group size on CSE effectiveness has not received a lot of attention in past research. In particular, we are aware of only one work that focuses on this issue: Joho et al. [70], who performed elaborate simulations to investigate the effect of group size changes in recall-oriented search tasks. Simulations, though, are limited in their ability to model the real world, and thus, in this chapter, we conducted a user study to investigate to what extent the findings of this simulation study hold in a setup with actual users. In particular, we aimed to answer the research question:

RQ1. What is the impact of group size on CSE in terms of retrieval effectiveness and search behaviours?

To this end, we designed a crowd-sourcing based user experiment using our CSE framework `SearchX` capable of synchronising crowd-workers across tasks and user groups to explore the extent to which the simulation results hold in practice. In particular, of the six hypotheses we investigated (all focusing on the impact of group size changes on search effectiveness), we find partial support for only two of them (**H1** and **H4**), demonstrating ultimately the limitations that a simulation setup suffers from. In addition, our preliminary results show evidence to support **H5**.

Specifically, we do not observe diminishing returns with increasing group sizes; the group recall steadily increases as more collaborators participate in the search. We also do not find large collaborating groups to be most beneficial at the start of a search session. Instead, the increased recall obtained early in the search session in contrast to smaller collaborating groups is retained throughout the search session. Also, we noted that our results also confirm our intuition that sharing of knowledge approaches—which work well in simulations as they assume perfect relevance judgement capabilities and no increased cognitive load with increasing group size—need to be considered with care as group sizes increase. Lastly, our results on search space exploration show preliminary evidence that larger groups are more effective in discovering different areas of the search space that single users would not discover by themselves.

We believe that the results we presented in this chapter are an important step towards a greater research emphasis on changing and importantly *increasing* group sizes in CSE settings. In particular, in our experiments, we considered a maximum of six collaborators due to inherent limitations that the mix of synchronous CSE and crowd-working platforms suffer from, however, at the same time, we did not find evidence that we already reached the upper recall bound with our maximum group size. Our findings show that there is room to investigate how effective are CSE systems to support groups of size beyond 2-3, and it is pertinent to consider other group sizes for future research.

Based on our findings, several interesting avenues of future work can be explored:

- We believe, based on our experiences in this chapter, we have reached close to the maximum number of collaborators we can engage synchronously in a crowd-sourcing setup. This is because to form groups larger than six people, one would need to increase the waiting time to form a crowd-sourcing group and the willingness of participants to wait. Consequently, this would increase costs and time for such experiment. As a next step, future work should investigate the deployment of a CSE system in a large-scale environment where larger groups than six work together towards a shared goal.
- One future avenue is to investigate the modelling of the CSE process via the recently introduced economic models of search [10] in order to gain a better theoretical understanding of the interface and algorithmic mediation approaches that are worth exploring further. For instance, such theoretical models could help us identify which collaborative feature would be most beneficial for collaborative groups before conducting a user study experiment.

4

Analyzing the Effects of Role Assignment in CSE Tasks

In the previous chapter, we considered group sizes as one aspect of group configuration in CSE. In this chapter, we explore another aspect of group configuration, assigned roles *in CSE: instead of groups determining how to split up the search task (and thus losing time communicating and coordinating), group members are assigned different roles by the CSE system. This is hypothesised by previous works to make the search process more efficient and more successful. Past user studies have provided contradicting evidence as to the utility of assigned roles in CSE. This chapter aims to settle the question of the effectiveness of role-based CSE by conducting a user study across four conditions. Apart from search effectiveness, we also investigate the differences in search behaviour and communication patterns between those conditions. We find that search effectiveness does not differ significantly between groups with and without assigned roles, even though our search behaviour and communication pattern analyses reveal a considerable divergence among those groups. We thus have to conclude that role-based CSE—at least when employing the roles as established in the literature—is not beneficial for CSE. Users are just as effective when self-organising as a group.*

4.1 Introduction

Past research efforts in CSE have focused on both the search backend [129, 135, 136] and frontend [41, 102, 109] to improve the search experience of collaborating searchers. Intuitively, searchers who independently work towards solving a shared task perform duplicate work, which can be avoided by explicit collaboration when *communication* and *coordination* of the search effort are available. Apart from user interface and retrieval algorithms' improvements, prior works have also considered the impact of different group attributes, including the size of the group [96], the expertise of the group members [136], the nature of the collaboration (synchronous vs. asynchronous, co-located vs. remote [32, 109]) and the *roles* searchers take up in a group [111, 129, 144]. The latter is the focus of this chapter: when searchers are collaborating, they can either communicate to decide how to share the work (e.g., by focusing on different aspects of the information need) or be assigned pre-defined *roles*. These roles specify each group member's function (e.g., one group member is responsible for finding documents, another one for reading and evaluating them). This in turn, in previous works is hypothesised to reduce the amount of communication necessary, thus leaving searchers with more time to focus on the task at hand[111, 129, 144]. Ultimately, this should result in better task outcomes.

Despite the intuitive appeal of role-based CSE (which in our case in this chapter and that of prior work is always restricted to *pairs* of searchers), prior user studies have revealed conflicting evidence. While [111] showed that groups with pre-assigned roles outperform artificial groups (i.e. groups created from the log traces of individual searchers), [144] found that groups without pre-assigned roles tended to reach a higher search effectiveness than groups with assigned roles. And lastly, [69] did not observe significant differences in retrieval performance between groups without pre-assigned roles and artificial groups. Table 4.1 shows that groups with pre-assigned roles have not definitively been shown to offer improved precision and/or recall over groups without pre-assigned roles. In summary, from Table 4.1, we find role-based CSE to be both better and worse than CSE without pre-assigned roles: $\mathbf{COLR} \overset{\text{row 2}}{>} (\mathbf{COL} \overset{\text{row 1}}{\approx} \mathbf{INDM}) \overset{\text{row 3}}{>} \mathbf{COLR}$. We used the sign ≈ to mean approximately the same.

Table 4.1: Overview of previous works comparing different CSE settings in a user study setup. Prior major findings in terms of retrieval effectiveness are listed in the last two columns. Four types of settings are considered: a pair of collaborating users (**COL**); a pair of collaborating users with pre-assigned **R**oles (**COLR**); an individual searcher (**IND**); and an artificial pair of users (**INDM**), created by the **M**erging of two individual searchers' log traces.

	COL	**COLR**	**IND**	**INDM**	**Precision**	**Recall**
	👤 ↔ 👤	👤 ↔ 👤	👤	👤 + 👤		
Joho et al. [69]	✔			✔	≈	≈
Pickens et al. [111]		✔		✔	**COLR>INDM**	**COLR>INDM**
Tamine and Soulier [144]	✔	✔			**COL>COLR**	≈
This study	✔	✔	✔	✔		

The major limitation of these prior works is the binary nature of the comparison: in each case, only two conditions were compared. In this chapter, we aim to settle the question of the effectiveness of *role-based* CSE by conducting a user study across four conditions: (i) individual searchers; (ii) artificial groups created by aggregating the search logs of pairs of individual searchers; (iii) pairs of collaborating searchers without assigned roles; and (iv) pairs of collaborating searchers with assigned roles. We do not only consider search performance, but also explore the differences in search behaviour across the four conditions as well as the impact of roles on searchers' communication patterns. In this chapter, we aim to answer the following main research question:

RQ2 What are the benefits of role-based CSE in recall-oriented tasks in terms of search retrieval effectiveness, search behaviours, and communication patterns?

More specifically, we will answer **RQ2** with the following three fine-grained research questions:

RQ2.1 Does retrieval effectiveness between collaborative groups with unassigned and pre-assigned roles differ?

RQ2.2 Given the collaborative groups outlined in **RQ2.1**, how do search behaviours differ between them, and compared to searchers working individually?

RQ2.3 To what extent does the assignment of roles affect communication between group members?

Based on our crowd-sourced user study (120 participants) with a simulated work-task setting [27, 44] and three search topics, our main findings in this chapter can be summarised as follows:

- Precision and recall do not differ significantly between groups with and without assigned roles. In addition, artificial groups (i.e. pairs of individual searchers) who do not benefit from communication and coordination perform at the same level of recall as collaborative groups.
- Groups with assigned roles exhibit search behaviours that indicate their compliance with the given role distribution: group members tasked with the exploration of the search space spend less time reading documents, and issue more queries. In contrast, group members tasked with exploring the set of documents pre-selected by other group member do indeed spend more time reading documents. Despite the apparent differences in search behaviours, search performance overall does not differ, as indicated above.
- In terms of communication patterns, we find groups with assigned roles communicate less frequently than groups without assigned roles. However, the time saved from fewer exchanges does not translate into an increased level of search performance.

4.2 Related Work

In Chapter 2 we have introduced related work on CSE systems, and SearchX, our CSE system. In this section, we introduce related work focus on role-based aspects of CSE.

Roles are important within any team—unsurprisingly, there has been a rich line of research examining this phenomenon within CSE. Shah et al. [129] proposed that a CSE user's role is either *task-based* or *information content-based*. In task-based roles, each user within a group is responsible for a certain kind of task/activity. In contrast, information content-based roles involve all users of a group having the same tasks, but are served different search results based on the tasks they are assigned. Indeed, role-level algorithmic mediation research dates back more than a decade [63, 111, 129, 135, 136, 138, 144].

Task Based Roles

On the other hand, concrete examples of task-based roles are the roles of *prospector* and *miner* [111]. The prospector is primarily responsible for querying the system and identifying a series of potentially useful documents, whereas the miner is responsible for 'drilling down' into the documents, expending more effort in analyzing the said documents for usefulness/relevance.

To investigate if mediated collaboration offers a more effective means of searching than a post-hoc merging of independently produced results (from independent searchers who are unaware of the other's presence), Pickens et al. [111] conducted a user study with system-driven mediation. Their working task was focused on multimedia retrieval with a search engine specially adapted for retrieving video shots. As such, the collaborative mediation was supported with two different UIs—one for the prospector (with algorithmic mediation), and one for the miner. While each participant had their own interface, both had a common display to view the final list of selected shots. Comparing this collaborative approach against a *merged* approach, where two independent searchers' results were grouped together, the authors found that the collaborative search setup offered better performance than post-hoc merging. The collaborative groups also found more unique, relevant video shots earlier in the search sessions.

These findings suggest that CSE teams find more unique, relevant documents than merged results from two independent searchers. Indeed, the advantage that CSE teams exhibit appears to be larger over topics with a higher degree of difficulty—and that for these topics, there is a more significant advantage to be had *earlier* in the search session.

Information Content-Based Roles

An example of information content-based roles are the *gatherer* and *surveyor* roles [129]. The gatherer's search results are retrieved in a way that they can broadly explore the information landscape (hinting towards the *diversification* of search results), while the surveyor's search results are tailored towards results that focus on a specific aspect of the topic being considered. The authors described how their proposed technique is able to provided a list of high-precision results required by the gatherer role and a list of diverse search results, which is required by the surveyor. Their focus was to evaluate the method rather than evaluate the proposed roles of gatherer and surveyor.

A further example of information content-based roles are the *domain expert* and *domain novice* as proposed by Soulier et al. [136]. Here, document rankings and allocation

Table 4.2: Overview of key statistics of empirical evaluations of role-based collaborative search. From left to right, **Roles** outlines the different searcher roles trialed. *Study Type (ST)* considers: *[Sim. Ind.]*, a simulation study with batch evaluation using *single-search logs*; *[Sim. Col.]*, a simulation study with batch evaluation using *collaborative search logs*; and *[User study]* represents user study in a lab or crowdsourced environment. **Division of Labor (DL)** considers: *user-driven*, *system-driven*, or *user-driven system mediation*. Group sizes are summarized by the number of groups, **#G**. The number of search tasks per group is denoted by **#T**. Work task type (**WT**) is also provided, and finally, **Collection** refers to the data collection used. For contrast, the study reported in this paper is also included in the last row of this table. Note that any information missing in the respective paper is indicated with –.

Study	Roles	ST	DL	#G	#T	WT	Collection
Pickens et al. [111]	Prospector/Miner	User study	System	4	24	Multimedia	TRECVid07
Shah et al. [129]	Gatherer/Surveyor	Sim. Ind.	System	5	10	–	–
Soulier et al. [136]	Domain expert/novice	Sim. Ind.	System	–	20	–	TREC Vol. 4
Htun et al. [63]	Non-uniform access	Sim. Ind.	System	55	13	–	AQUAINT
Soulier et al. [135]	Mined [111, 129] roles	Sim. Col.	User-system	70	2	–	Web
Tamine and Soulier [144]	[111, 129] roles	User study	User	75	1	Recall oriented	Web
Soulier et al. [138]	Latent roles	Sim. Col.	User system	2	70	–	Web
This study	**Prospector/Miner**	**User study**	**User**	**48**	**3**	**Recall-oriented**	**AQUAINT**

are adapted based on user expertise. Works by Shah et al. [129] and Soulier et al. [136] compare algorithmic mediation approaches that are optimized based on roles. Shah et al. [129] found that role-based algorithmic mediation led to a higher degree of search effectiveness than the naïve merging of search results, again undertaken by independent pairs of searchers. Similarly, Soulier et al. [136] found that role-based algorithmic mediation was more effective than simply providing searchers with results using an ad-hoc ranking function based on BM25.

Determining Roles

Researchers have also attempted to determine what role a user should be allocated based on attributes such as their search history. Soulier et al. [135] proposed a user-driven system mediation technique to identify pre-defined roles using collaborative search logs from a previous study undertaken by Tamine and Soulier [144]. They later utilised the identified roles to personalise rankings. In this study, participants were pre-assigned roles (e.g., either a prospector or miner), or were not assigned roles at all. The authors report that their technique to mine roles from collaborative search logs was more effective for groups that were pre-assigned roles versus not assigned roles. They also showed in a simulation study that optimising search results based on roles improves retrieval effectiveness over non-role based algorithmic mediation.

Later, Soulier et al. [138] proposed an approach to mine roles by identifying pairs of latent features. Features, for example, include the number of queries issues, and the number of clicks that were made on a *Search Engine Results Page (SERP)*. These features could be combined together to provide a search behaviour *'pattern'* to identify what role the participant was more suited to. The identification of the pattern could then be used to determine how to optimize the results for each participant. The technique showed improvements over the previous study undertook by Soulier et al. [135], which used the pre-assigned prospector and miner roles.

In order to understand the different effects of assigning roles to users in a group, Tamine and Soulier [144] conducted a user study comparing how assigning roles from the studies by Pickens et al. [111] and Shah et al. [129] and letting individuals work freely to their own (i.e., without pre-assigned roles) brief impacted on collaborative search performance. They found that when working without pre-assigned roles, participants worked less independently, and were likely to broaden their responsibilities for reaching a goal. However, roles did emerge and were shown to happen early in the search session. Conversely, those who were assigned to roles sometimes found it difficult to *'toe the line'* with their allocated brief.

Summary

Given the complexity and range of different studies (under different experimental permutations) in the CSE domain, we provide a summary table of the main studies in the area of previous works [63, 111, 129, 135, 136, 138, 144].

Table 4.2 provides the summary over several key aspects in this area (such as what roles were used, what the type of experiment was, etc.)—as well as providing an entry in the table for the present study. In Table 4.2, we find the majority of previous works to have been based on simulation studies. We can see as well that even though other roles have been studied in the literature, the prospector/miner roles were focused of many studies, including this chapter [111, 144].

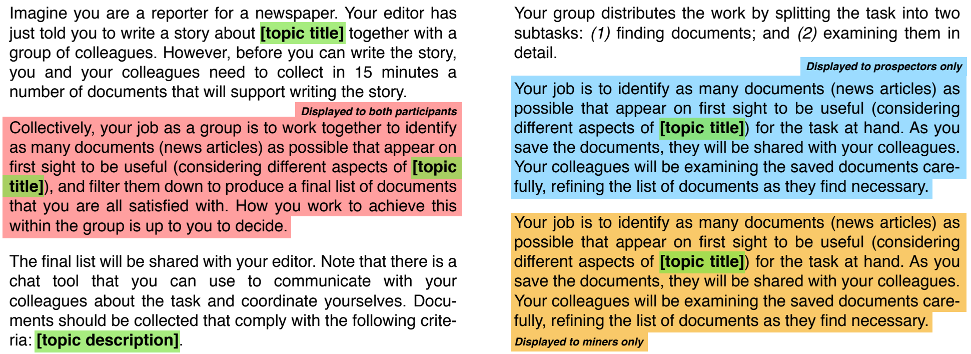
Other roles might also be defined based on what people are responsible for in the context of searching in the enterprise. For instance, someone in sales and someone in R&D working on the same task would presumably have different roles searching together as they have different roles in the company. For instance, roles in a company may be identified given psychological traits or team roles (e.g., Belbin Team Roles [17]). The goal of this research is to complement existing works that have examined role-based CSE in which users are assigned a pre-determined role in the literature. Given the summaries of prior works in tables 4.1 and 4.2, we aim to examine in this chapter how the assignment of roles impact retrieval effectiveness and search behaviours (**COLR**) over a number of different topics, versus when they are not present (**COL**)—and compare this to a baseline where searchers work independently, and through a simulated group by merging two independent searchers' activities together (**INDM**).

4.3 Experimental Design

In this section, we turn our attention to detail in our experimental design. In order to address our research questions, we ran a between-subjects crowd-sourced user study, where participants were assigned to conditions that either tasked them to search individually or work with another participant in a collaborative fashion. Participants were required to work over three individual tasks.

4.3.1 SearchX Setup

For this experiment, we utilise SearchX framework described in Chapter 2. Specifically, our participants were provided SearchX V3 with the following features (the basic UI interface can be seen in Figure 2.3). For single-user search participants, we provided SearchX with a search box without search verticals ① along with collection-based query autocompletion (see Figure 2.4b), a document viewer (see Figure 2.8) without document rating ⑨



Imagine you are a reporter for a newspaper. Your editor has just told you to write a story about **[topic title]** together with a group of colleagues. However, before you can write the story, you and your colleagues need to collect in 15 minutes a number of documents that will support writing the story.

Displayed to both participants

Collectively, your job as a group is to work together to identify as many documents (news articles) as possible that appear on first sight to be useful (considering different aspects of **[topic title]**), and filter them down to produce a final list of documents that you are all satisfied with. How you work to achieve this within the group is up to you to decide.

The final list will be shared with your editor. Note that there is a chat tool that you can use to communicate with your colleagues about the task and coordinate yourselves. Documents should be collected that comply with the following criteria: **[topic description]**.

(a) Task template for condition COL

Your group distributes the work by splitting the task into two subtasks: *(1)* finding documents; and *(2)* examining them in detail.

Displayed to prospectors only

Your job is to identify as many documents (news articles) as possible that appear on first sight to be useful (considering different aspects of **[topic title]**) for the task at hand. As you save the documents, they will be shared with your colleagues. Your colleagues will be examining the saved documents carefully, refining the list of documents as they find necessary.

Your job is to identify as many documents (news articles) as possible that appear on first sight to be useful (considering different aspects of **[topic title]**) for the task at hand. As you save the documents, they will be shared with your colleagues. Your colleagues will be examining the saved documents carefully, refining the list of documents as they find necessary.

Displayed to miners only

(b) Task template for condition COLR

Figure 4.1: Task information templates presented to participants, for both **COL** *(a)* and **COLR** *(b)*. Text highlighted in green is replaced with details of the topic considered; text in red is shown to both participants (for **COL** only). Text in blue outlines the role for a prospector, with text in orange outlining the role for a miner. The final two are shown in condition **COLR** only.

Table 4.3: Overview of the four experimental conditions.

Collaborative Conditions	
COL	CSEers *without role assignment*
COLR	CSE users *with explicit role assignment*

Independent Conditions	
IND	Independent search with individual lists of saved documents and issued queries
INDM	Merging sessions of independent searchers to produce a *simulated collaborative group*, with individual lists of saved documents and issued queries *merged* together (taking the union, $\bigcup$, of the two lists)

and document annotation ⑩, recent queries ④, and saved documents ⑤. For the synchronous CSE we also included the interface elements recent queries ④ and saved documents ⑤ with anonymous icons for each group member (see Figure 2.6b in Chapter 2), group chat ⑥ with our own implementation (Figure 2.5b), and a users' status bar ⑦.

4.3.2 Experimental Conditions

In our motivation, we outlined a number of experimental conditions that have been trialled in previous works (summarized in Table 4.1). For this study, we draw on all four previously defined conditions, as in Table 4.3. The four conditions consider searching under both collaborative and individual contexts, with the added caveat of preassigned roles for the CSE groups.

We explored two collaborative experimental conditions; as shown in Table 4.3, these are **COL** and **COLR**. While both collaborative in nature, **COL** did not enforce preassigned

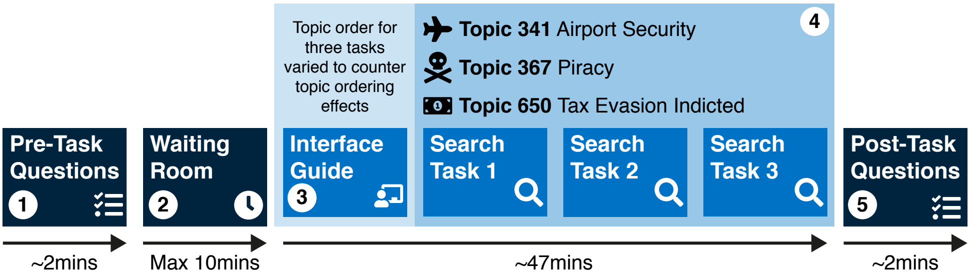


Figure 4.2: The experiment workflow for participants of this study. For more detail on the five main steps (as numbered on the diagram), refer to Section 4.3.5.

division of labour for each participant, treating both participants of a group symmetrically. For **COLR**, participants were treated asymmetrically—roles were preassigned. This allowed us to examine how assigning roles to participants of CSE groups impacts retrieval effectiveness and search behaviours. Both of these conditions were drawn from the literature—**COL** corresponds to the *w/o roles* condition as used by [144], while **COLR** corresponds to the experimental condition *PM* also used by [144], along with the *Collab.* condition as used by [129] (albeit without algorithmic mediation). For condition **COLR**, we assigned participants to act as either the *prospector* or *miner*. These two roles have been studied in many previous works [111, 129, 138, 144, 144]. We chose these two roles to enable us to compare our findings against two seminal studies in this area [111, 144]. Other possibilities such as giving participants the freedom to choose one of the roles could have been addressed. However, our research questions are focused on preassigned roles. Hence, we leave these possibilities for future work.

Participants for each collaborative condition were provided with a predetermined task outline, providing details on what exactly their role in the search tasks would entail. The task templates we devised are illustrated in Figure 4.1. On the left (subfigure *(a)*), the template for symmetric participants (under condition **COL**) is shown (in red). On the right (subfigure *(b)*), the template is shown for both prospectors (in blue) and miners (in orange). Only the relevant instructional text block would be shown for individual participants. Note that topic-specific information replaces the filler text highlighted in green.

In contrast to the collaborative conditions, we also considered experimental conditions where participants were treated as independent searchers (i.e. searching alone, without awareness of others, or any division of labour). This condition, **IND**, served as our baseline to which we could compare the collaborative conditions in our analysis. Under this condition, participants also used the `SearchX` interface, without collaborative features (such as 6 and 7, as per Figure 2.3. Participants were instructed as per the task template for **COL** (Figure 4.1 *(a)*), but without the inclusion of the red block.

As per Shah et al. [129], we also employed a fourth experimental condition, where pairs of independent searchers (**IND**) were combined to form a *simulated group*, with the data recorded for their respective search sessions *merged* together. This meant, for example, that the list of saved documents for a simulated, merged group would be constructed from the two saved document lists of independent searchers. This additional merged condition, **INDM**, allowed us also to explore the extent of the simulation of CSE compared to actual collaborative workers (i.e., from **COL** and **COLR**).

4.3.3 Retrieval Backend and Topic Selection

In this experiment, we used TREC AQUAINT corpus. Here our retrieval backend used language modelling, details are described in Chapter 2.3.4. We used the same topics selected in Chapter 3. They were *airport security*, *piracy*; and *tax evasion indicted*—details and statistics for each topic are provided in Table 3.3.

4.3.4 Crowdsourced Participants and Allocation

Although a majority of CSE research has been conducted in a lab-based setting in which participants know each other, this study reports on a crowd-sourced experiment. As the experiment is crowd-sourced, it is highly unlikely that participants know whom they work with. Although this is a potential limitation of our study, we argue that we imposed sufficient controls on potential recruits for this study to ensure a pool of high quality workers. Besides, the CSE task that we designed in this chapter is a work-task. This type of task does not require participants to know about each other differently from CSE tasks such as decision-making tasks (e.g., travel planning and online shopping). In addition, crowdsourced, collaborative tasks are becoming more commonplace in contemporary research [66, 81, 117, 152] as well as in CSE studies [95, 96, 163].

Our experiment was undertaken on the *Prolific* platform, which has been shown to be a more reliable crowd-sourcing platform for cognitively demanding tasks than FigureEight and MTurk [110]. Over the course of seven days in April 2020, a total of 120 participants were recruited from the platform. To ensure high quality workers and reduce noise in our collected data, we required that participants had at least a 90% approval rating on the platform, and had completed at least 50 experiments on Prolific. Their median age was 33 (min: 18, max: 69). Of the 120 participants, 69 participants self-identified as female, with 51 self-identified as male. Most participants (74) were from the UK, with the remaining participants from English-speaking countries such as Australia, the United States, Ireland, and Canada. Their academic backgrounds varied: 47 reported a high-school diploma or lower as highest academic degree, 51 a bachelor's degree, and the remaining 16 a graduate degree or six had other types of degree. Participants were recompensed with £6.50 per hour for the experiment, which was above the price suggested by the Prolific platform. Finally, we ensured that participants could only participate in the experiment once.

We randomly allocated 24 participants to the **IND** condition. For collaborative conditions **COL** and **COLR**, we allocated 48 participants to each condition, meaning 24 groups for **COL** and 24 groups for **COLR**. Groups of two participants were formed as the experiment was started. Three search tasks, each task a topic, were undertaken during the course of the experiment; topic orderings were determined via a Latin-square experimental design to mitigate topic ordering effects [72]. [144] employed a different topic per condition. This means that the interaction between topic and condition can have a confounding effect, in which both topic and condition impact the results. Our experimental design addresses this issue, allowing for a fair comparison between topics and conditions.

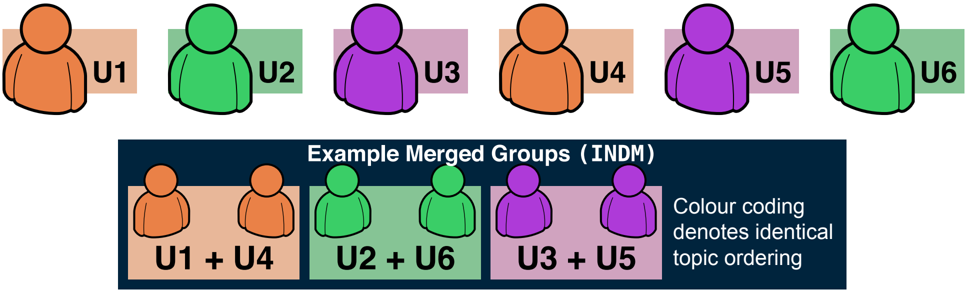

U1
U2
U3
U4
U5
U6
Example Merged Groups (INDM)
U1 + U4
U2 + U6
U3 + U5
Colour coding denotes identical topic ordering

Figure 4.3: Example illustration demonstrating how merged groups for condition **INDM** were created over six sample participants, U_1 to U_6. Different colours denote a different ordering of topics—only those who completed each topic in the same order were merged. The black box at the bottom shows the three merged groups.

Merging Independent Participants

Topic ordering is an important consideration when creating simulated, collaborative groups—or *merged groups* (**INDM**). In order for us to merge independent participants together, we consider participants that were given the same topic ordering. Only participants with the same topic ordering assignments were merged together, leading to a total of 36 simulated collaborative groups based on the 24 participants allocated to the **IND** condition. In Figure 4.3, we depict how we merged independent participants. As a concrete example, **U1** and **U4** have the same topic orderings. Therefore we created an artificial group merging these two users. As discussed previously, merged results bring together behavioural and interactions of two distinct users. We merge the saved documents, issued queries, search behaviours (such as clicks and time spent reading documents) of two users together to form a group.

4.3.5 Experimental Procedure

Once participants had accepted the task on the Prolific platform, they were redirected to our server setup with the instance of `SearchX` running for their assigned condition. A broad overview of the experimental procedure is shown in Figure 4.2. We outline each of the main steps as follow.

1. The experiment begins with the completion of a short pre-task questionnaire.

2. After completion of the initial questionnaire, participants in the collaborative condition were then moved to a *waiting room*, where they waited for other participants to arrive. A ten-minute limit was imposed; we provided a countdown timer and a game of Tetris for participants to play while they waited. The option of an audible alert was also provided when a second participant joined.

3. Once the participants moved to the CSE phase, they first undertook a short interactive guide of the `SearchX` interface. The key features of the interface were highlighted to the participants, along with a short explanation of what the highlighted components do.

❹ After the interface guide, work began on the three search tasks. Both participants undertook the same topic at the same time; topic ordering was determined on a per-group basis, not on an individual basis. After 15 minutes of searching, the session would end, and participants were moved to the next search session. For a fair comparison, we chose 15 minutes per topic, as previous work also chose this duration [69, 111]. The process would repeat a total of three times. This was all done in sync to ensure that participants would begin searching at the same point in time.

❺ Once all three search activities were completed, we then provided a final, post-task questionnaire. The details of the questions we posed can be found below. Upon completion of this questionnaire, the participants were returned to the Prolific platform.

For the independent condition (**IND**), the waiting room was obviously skipped—no waiting was required for another participant to begin. This meant that after step ❶, the participant moved straight to step ❸. We also considered the scenario where a participant, allocated to one of the collaborative conditions (i.e. **COL** or **COLR**), found themselves waiting for 10 minutes in the waiting room. The waiting room contained a warning message that stated if the 10 minute period elapsed without anyone else joining, they would then be switched to the **IND** condition. From that point on, task templates were switched to match the **IND** condition, and all collaborative features (e.g., the chat component) were removed from the `SearchX` interface.

To provide context for the ad-hoc search topics, we employed the task templates as shown in Figure 4.1—the exact template shown depended on the condition that participants were assigned to. These recall-orientated tasks can be found in scenarios such as patent retrieval and e-discovery, and represent a typical task in CSE which spurs the required complex and exploratory search behaviours, as discussed by [101]. For the independent condition, we did not include the information about a *group of colleagues* searching together, and we did not include the description of the chat tool. To maintain consistency among participants, we employed for all participants the same pre-task questionnaire and post-task questionnaire.

4.3.6 Pre-Task Questionnaire

Our pre-task questionnaire, besides demographic questions, included six CSE questions as presented by [101] to aid priming of the participants for the upcoming CSE tasks. We included a short explanation of CSE and when and how it can happen, along with three images of groups of people collaborating (co-located and remote collaboration). We began by asking two questions about CSE episodes:

- Have you ever collaborated with other people to search the Web? *[Yes, No]*
- How often do you engage in collaborative Web search? *[Number]*

After these questions had been answered, we asked additional questions:

- Describe what you were looking for. For example, you may be a husband and wife planning a trip for your family, a group of students working on a writing assignment and sharing search results/findings, or a couple shopping for a new sofa. *[Text]*

- How many others did you collaborate with (not including yourself)? *[Number]*
- Given the example scenario you provided above, describe how you and your partner, friend or colleague performed the search activity. As examples of what we are interested in, did one person perform all the exploratory searching, with another examining the findings in detail? Or did each person involved look after individual components of the search activity (e.g., when booking a holiday, did one person focus on booking flights, with another looking at a hotel)? *[Text]*
- When performing the example scenario you provided, how did you communicate with others you were performing the search activity with? *[Text]*

4.3.7 Post-Task Questionnaire

As previously mentioned, we provided a post-task questionnaire to capture each participant's *experiences* of working collaboratively with the provided system. We began by asking two questions with simple responses to gauge their overall impressions of the experiment.

- How many people did you just now collaborate with (not including yourself)? *[Number]*
- It was easy to determine if a document was relevant to a task. *[Disagree, Agree]*

After these questions had been answered, we asked eight additional questions. Answers were provided with a 5-point Likert scale, where the focus of the questions ranged from how they thought the system worked, to how well they got on with their search partner.

- It was easy to understand why documents were retrieved in response to my queries. *[Disagree, Agree]*
- I didn't notice any inconsistencies when I used the system. *[Disagree, Agree]*
- How difficult was this task? *[Very easy, Very difficulty]*
- It was easy to understand what was my job in the task. *[Disagree, Agree]*
- I could communicate well with my partners. *[Disagree, Agree]*
- I could see what the others were searching for and were examining during the task. *[Disagree, Agree]*
- It was easy to share what I found useful during my searches with my partners. *[Disagree, Agree]*
- Did you find the collaborative features useful (multi-grid question with one row for each feature: recent queries, saved documents, chat tool, and hiding saved and excluded results)? *[Disagree, Agree]*

4.3.8 Evaluation Measures

In order for us to be able to address our three research questions, we relied on a range of evaluation measures that are commonly employed measures in CSE research [69, 129, 136, 145]. We consider three explicit categories of measures: *retrieval effectiveness*; *search behaviours*; and *conversational measures*.

Retrieval Effectiveness Measures

To measure retrieval effectiveness, we employ *group recall (GR)* (already employed in Chapter 3 in Equation 3.1) and *group precision (GP)* which is defined as the recall and precision for the union of the sets of documents that each participant in a collaborative group saved. Or, expressed formally,

$$GR(g,t) = \frac{|\bigcup_{p \in g} rel\_saved(g,t,p)|}{|reldocs(t)|}, \tag{4.1}$$

and

$$GP(g,t) = \frac{|\bigcup_{p \in g} rel\_saved(g,t,p)|}{|\bigcup_{p \in g} saved(g,t,p)|}, \tag{4.2}$$

where g represents a group, t represents a topic, and p represents a given participant. Function rel_saved returns the saved documents that are TREC relevant, for a given g, t and u. $reldocs$ returns the TREC relevant documents for topic t, while $saved$ returns all of the saved documents for a given g, t and p. To calculate the average group recall $\bar{G}R(c,t)$, and average group precision $\bar{G}P(c,t)$ for all groups in an experimental condition c and topic t, group recall and precision is averaged across all groups in an experimental condition.

We also measured *individual precision* and *individual recall* for each participant in a group for the collaborative conditions, similarly to *group recall* and *group precision*, however, we consider the saved documents by participants individually. In our analysis, for the independent condition, we measured precision and recall individually as we only have one participant per group.

Search Behaviour Measures

We also report on a number of aggregate search behaviours. We report aggregate search behaviours by group and individual. Included in our reporting are a number of behaviours associated with querying, including:

- the mean number of queries issued;
- the number of unique queries (i.e., exact same queries);
- the average query length (in the number of terms used);
- the number of unique query terms used (i.e. we split query terms for all issued queries without applying stemming); and
- the number of clicks on previously issued queries (from interface component 4 in Figure 2.3 in Chapter 2).

We considered interactions with the wider SERP, considering events involving clicks on documents, for example.

- the number of clicks on results on SERPs;
- the number of unique clicks on results on SERPs;
- total number of clicks on documents that *were* TREC relevant;
- the number of unique snippets of documents viewed on SERPs (i.e. the number of unique search results viewed on SERPs); and
- the number of unique snippets of documents viewed on SERPs that *were* TREC relevant;

We also recorded the amount of time that was spent reading documents *(document dwell time)*, allowing us to calculate the total document dwell time, and the average document dwell time.

Finally, we also logged events pertaining to the overall *performance* of the participants (and therefore groups). This included a count of:

- the number of documents saved (i.e. *considered* relevant);
- the actual number of documents saved, that *were* TREC relevant; and
- the number of clicks on previously saved documents (from interface component ⑤ in Figure 2.3 in Chapter 2).

We report aggregate search behaviours by group and individual for our experimental conditions. For the group search behaviours, we sum up individual search behaviour measures (e.g., we sum up the number of queries by participants in a group).

Communication Pattern Measures

Our final set of evaluation measures considered the chat interface, or a series of measures that could be used to evaluate the chat messages exchanged between participants in conditions **COL** and **COLR**.

In order for us to be able to evaluate chat messages, a system of *message codes* was required. We adapted a sequence of codes as proposed by [163]—these codes have been acquired from previous collaborative studies [55, 102, 144] to evaluate the frequencies and temporal distributions of different types of chat messages. After a first inspection of the chat messages, we added two categories to this coding scheme specific to our task:

(CL) We noticed that many of our messages belonged neither to the TC or TD category (see Table 4.4). These messages discussed curating the list of documents (i.e. which documents to include or delete from the previously saved documents from interface component ⑥ in Figure 2.3). Therefore, we created a new category to describe this type of message.

Table 4.4: The coding scheme used for evaluating the exchanged messages between participants in experimental conditions **COL** and **COLR**. The schema is derived from the work by [163].

Code	Description	Examples
TC	Planning search strategies and division of labor	*"I will focus on US. Can someone else focus on UK?"* *"How would you like to divide this task?"*
TD	Discussing facts related to the general task topic	*"The Pavarotti one I think is in Berlin"* *"I am struggling to find any more for the US"*
CL	Discussing what documents to keep in the list	*"Should we whittle it down " "Delete it"*
SK	Sharing knowledge discovered during the search	*"the articles seem to focus on a smaller number of incidents overall "* *"I've looked for a few and am now looking at china as that seems to have had a fair amount of piracy"*
BR	Building relationships during the task	*"Hello there..." "good luck"*
AP	Appreciation of each others work	*"you did great!" "Good job. "*
ST	Technical issues of system being used	*"the search engine takes a while to load"* *"I thought it had crashed "*
OR	Others i.e. response or discussion not related to the topic	*"excuse the typos" "Think i'm done here"*

(ST) We observed that many chat messages in the (OR) category talked about system technical issues such as time to load search results. We created a new category only for this type of chat message.

The author of this thesis and two other contributors assessed the chat messages from this study. In total, 679 chat messages were assessed using the coding scheme and guidelines in Table 4.4. We randomly divided the chat messages into three equal shares. Each assessor was given a share, and 115 messages from the others share were added to compute an agreement score. In total, 163 chat messages (i.e. one-quarter of the messages) were overlapped among the three assessors, and the Fleiss' kappa inter-agreement score was 0.62, which indicates substantial agreement according to interpretations of the score in the literature. For the overlapped messages, we kept the majority voted category, and if a majority was not found, we assigned the others (OR) category. Similar to [55, 144], we computed these raw values and reported the proportions.

In addition, we computed communication patterns measures, namely:

- **Volume**, the number of messages exchanged by group members during the task;
- **Frequency**, the average time between messages exchanged between group members during the task;
- **Effort**, the number of times a group opened and closed the chat window; and
- **Balance**, the proportion of messages sent by an individual, collaborative group members.

Given the codes and the explanation on how these were applied to the recorded chat messages, a table with the codes, complete with example messages, is presented in Table 4.4.

Table 4.5: Topic-wise average group recall and group precision (averaged across all groups in a single topic/study-condition) across conditions averaged over groups. Statistical significance was computed independently for each topic; in each topic column, significant improvements ($p < 0.05$) are marked with superscript $^{\texttt{X}}$, and significant differences at $p < 0.01$ are marked with superscript underlined $^{\underline{\texttt{X}}}$, where X is the condition. **IND** figures are computed as *individual precision* and *individual recall* as we only have one participant per group.

	Group Precision			Group Recall		
Condition	**341**	**367**	**650**	**341**	**367**	**650**
IND	0.2748	0.7025	0.4569	0.0833	0.1085	0.1093
INDM	0.2417	0.7044	0.3326	$0.1293^{\underline{\texttt{IND}}}$	$0.1735^{\underline{\texttt{IND}}}$	$0.1550^{\underline{\texttt{IND}},\underline{\texttt{COL}},\underline{\texttt{COLR}}}$
COL	0.2923	0.7071	0.4319	$0.1315^{\texttt{IND}}$	$0.1820^{\underline{\texttt{IND}}}$	0.1076
COLR	$0.3452^{\texttt{INDM}}$	0.6365	0.4141	$0.1380^{\underline{\texttt{IND}}}$	$0.1728^{\underline{\texttt{IND}}}$	0.1163

4.4 Results

We now turn to our results, addressing each of our research questions in turn. For our analyses, we conducted for each topic a one-factor analysis of variance (ANOVA) with the experimental condition as our independent variable. We examined the ANOVA assumptions with Levene's test (homogeneity of variances) and Shapiro-Wilk's test (normality of the ANOVA residuals). We conducted a *post-hoc* analysis using Student t-tests and applied Bonferroni correction for multiple comparisons. Significant differences at $p < 0.05$ are marked with superscript, and significant differences at $p < 0.01$ are marked with superscript underlined $\underline{\text{X}}$. Along this section, we report p-values wherever a significant difference was found.

4.4.1 Retrieval Effectiveness

RQ2.1 considered *how retrieval effectiveness between collaborative groups with unassigned* (**COL**) *and preassigned* (**COLR**) *roles differ.* Table 4.5 summarises the retrieval effectiveness for group precision and group recall. We observe that group precision and recall do not differ significantly between groups with and without assigned roles. In terms of group precision, **COLR** groups significantly improve over **INDM** groups for topic 341. In terms of group recall, all the collaborative study conditions improve over **IND**, with the exception of topic 650. A possible explanation for the performance difference of topic 650 is that this topic leads participants to divide the task between the two subtopics (i) tax evasion in the U.S. or (ii) U.K. However, the TREC AQUAINT dataset has more relevant documents from the U.S. Therefore, the participants in the collaborative conditions **COL** and **COLR** may not perform as well as participants in the **IND** condition as they decided to look mostly at documents from the U.K. Consequently, the artificial groups formed by merging individual participants outperformed **COL** and **COLR** conditions for topic 650.

Let us now focus on individual precision and recall for the four roles we had in our study: individual (participants in condition **IND**), peer (participants in condition **COL**, i.e. without preassigned roles), prospector, and miner (both in **COLR**). From Figure 4.4, we see a contrast of the prospector and miner roles, where the prospector, on average, achieved a higher recall than the miner—but a lower precision. This is statistically significant for topic 367 ($p = 0.031$) in terms of recall for the miner when compared against the other

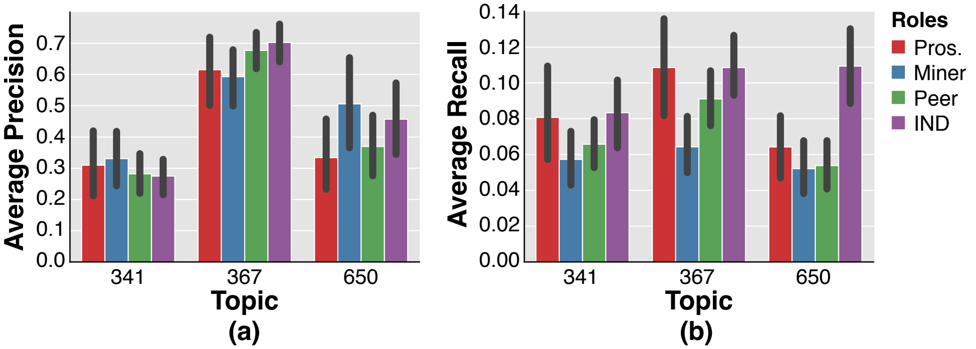


Figure 4.4: Overview of the mean precision and mean recall for each topic and condition across the four search roles trialled. **Left *(a)*:** average recall. **Right *(b)*:** average precision, where **Pros.** denotes *prospector*.

roles. Again, for topic 650, the single user role significantly outperforms collaborative roles ($p = 0.001$) in terms of overall recall. Although we chose similarly difficult topics, we also observe individual differences in performance for the three of them. This can be explained by the fact that participants may have found it easy to formulate queries for a certain topic, as we saw in some of the communication with other participants in the group.

To understand how precision and recall changed during search sessions, we performed a temporal analysis on group precision and recall along with the 15 minutes duration , over each of the three topics and four experimental conditions. Figures 4.5d-f show group recall over time. We can conclude from examining these plots that for some of the topics the collaborative groups performed better than independent searchers (**IND**) in terms of recall—again, except for topic 650. In terms of group precision, as shown in Figures 4.5a-c, we can observe that group precision peaks within the first few minutes of the search sessions (within the first three to five minutes). Interestingly, a drop off in precision is noted for topic 650 as the session progresses. A possible explanation is that participants could not find relevant documents for topic 650 (tax evasion indicted) as the collection contained fewer documents for one of the subtopics (tax evasion in the U.K.). This is in contrast to topics 341 and 367, where precision remains consistent (around 0.3 and 0.65 for topics 341 and 367, respectively). These results suggest that topics provide very different conclusions, which was also observed by Pickens et al. [111].

In summary, our analysis shows, overall, precision and recall do not differ significantly between groups with and without assigned roles. In contrast to prior works, we could not confirm the findings by Pickens et al. [111] (demonstrating improvements of **COLR** over **INDM**), and Tamine and Soulier [144] (demonstrating a greater precision for condition **COL** over **COLR**).

4.4.2 Collaborative Search Behaviour

Our second research question, **RQ2.2**, asked *how the search behaviours of collaborative searchers differed from those working independently.* Table 4.6 summarises search behaviours across both collaborative and independent conditions; here, we report the mean and stan-

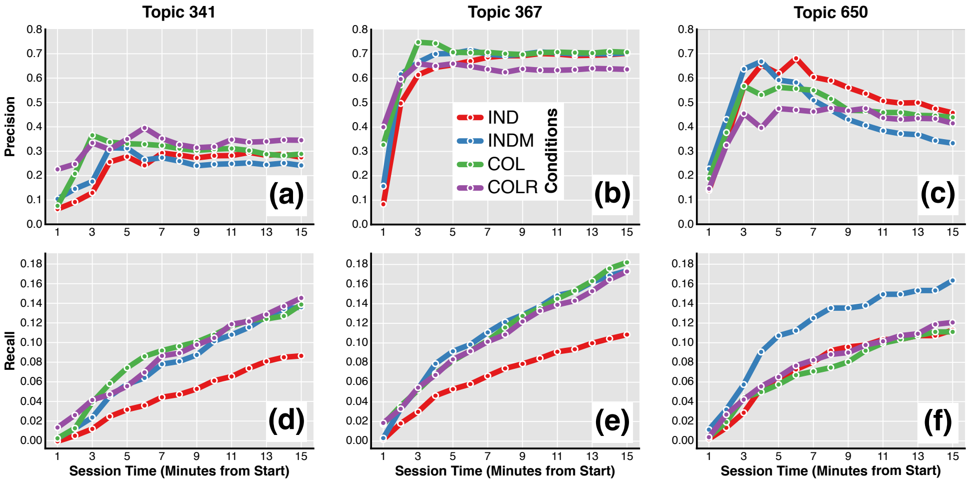


Figure 4.5: Overview of the average group precision (top row, *a-c*) and average group recall (bottom row, *d-f*) for each topic and search variant computed in one-minute time intervals.

dard deviation.

From Table 4.6, a number of observations can be made about the collaborative search behaviours. In terms of querying behaviour, comparing groups with assigned roles and groups without roles, we found that groups under the **COL** condition issue on average more queries (from 16.58 queries for **COL** to 14.50 queries for **COLR**), issue more unique queries (from 15.62 queries for **COL** to 12.72 queries for **COLR**, $p = 0.044$), and explore a larger set of query terms (from 24.11 terms for **COL** to 19.58 terms for **COLR**, $p = 0.032$). In terms of the click and saving documents behaviour, we observe that **COLR** groups have a similar click behaviour as **COL** groups (from 13.14 clicks for **COL** to 13.03 clicks for **COLR**) but spend more time reading documents (from 250 seconds for **COL** to 301 seconds for **COLR**). For the collaborative features' usage (interface components 4 and 5 in Figure 2.3), we find **COLR** groups to interact more with the *recent queries* (i.e. the query history) component (from 0.81 clicks for **COLR** to 0.26 clicks for **COL**, $p < 0.001$), and more with the *saved documents* component (from 6.58 clicks for **COLR** to 2.44 clicks for **COL**, $p < 0.001$). The observations we made here are expected as the guidelines we gave to **COLR** asked for the *prospector* to issue most of the queries during the search task, and the *miner* to read documents. Consequently, the interaction with collaborative features was higher for groups under the **COLR** condition than groups under the **COLR** condition. Finally, for the artificial groups **INDM**, we observe that, as expected, participants issued more queries and clicked on more results on the SERP than the **COLR** and **COL** groups. However, they spent less time reading and save fewer documents than both **COLR** and **COL**.

From these observations, similar to the findings by Tamine and Soulier [144], we can conclude that groups with assigned roles behave differently than groups without assigned roles, however, these differences in behaviour did not translate into improved retrieval effectiveness. In other words, assigning pre-defined roles did not overcome the overhead of

Table 4.6: Overview of collaborative and individual search behaviors across the study conditions. For **COL** and **COLR** conditions we depicted the values for individual behaviors. For each topic, the mean value is computed alongside standard deviation value; reported here is the average of those three mean values and standard deviation values.

	IND	**INDM**	**COL**		**COLR**		
			group	peer	group	prospector	miner
# Queries	9.67 (4.81)	19.33 (6.92)	16.58 (8.06)	8.29 (5.74)	14.50 (6.76)	7.76 (4.33)	6.74 (4.21)
# Unique Queries	9.23 (4.65)	18.29 (6.84)	15.62 (7.70)	7.93 (5.55)	12.72 (6.04)	7.04 (4.01)	6.18 (3.77)
Avg. Query Length	4.31 (2.10)	4.44 (2.29)	3.98 (1.40)	4.15 (1.96)	4.17 (1.38)	4.01 (1.66)	4.17 (1.59)
# Unique Query Terms	15.79 (7.72)	26.72 (9.08)	24.11 (14.20)	14.28 (10.30)	19.58 (8.14)	11.79 (6.06)	12.18 (6.99)
# Query History Clicks	0.10 (0.45)	0.19 (0.62)	0.26 (0.67)	0.13 (0.41)	0.81 (1.52)	0.18 (0.59)	0.62 (1.33)
# Clicks	9.36 (8.13)	18.72 (10.50)	13.14 (10.97)	6.57 (6.83)	13.03 (10.52)	6.81 (7.80)	6.22 (5.63)
# Unique Clicks	8.90 (7.73)	16.50 (9.05)	12.00 (9.69)	6.28 (6.53)	12.01 (9.04)	6.56 (7.45)	5.88 (5.16)
# Unique Rel. Clicks	2.94 (3.90)	5.25 (5.11)	3.64 (3.82)	1.89 (2.58)	3.86 (3.89)	2.08 (2.98)	1.94 (2.18)
# Unique Snippets	114.46 (53.55)	195.17 (70.69)	164.29 (63.70)	96.72 (50.06)	161.40 (56.94)	114.92 (53.45)	79.06 (35.66)
# Unique Rel. Snippets	11.53 (8.36)	17.14 (11.75)	15.36 (11.17)	10.39 (8.38)	14.88 (9.63)	12.06 (8.83)	8.71 (6.88)
Avg. Doc. Dwelltime	14.32 (15.09)	14.32 (10.87)	15.36 (11.17)	11.08 (11.62)	14.88 (9.63)	10.42 (9.21)	15.88 (13.87)
Total Doc. Dwelltime	162.40 (154.36)	324.81 (220.21)	250.61 (207.15)	125.31 (141.34)	301.53 (194.60)	110.90 (120.86)	190.62 (150.85)
# Saved Docs.	10.51 (6.08)	17.90 (7.93)	15.39 (9.15)	7.69 (5.58)	16.65 (15.25)	10.83 (13.14)	5.82 (4.45)
# Saved Rel. Docs.	4.78 (3.79)	7.44 (5.52)	7.31 (6.56)	3.65 (3.95)	7.39 (6.52)	4.50 (4.81)	2.89 (2.84)
# Saved Docs. Clicks	1.42 (4.00)	2.83 (6.26)	2.44 (5.09)	1.22 (3.11)	6.58 (7.29)	0.86 (2.35)	5.72 (6.22)

interacting with collaborative features such as previously issued queries and saved documents.

We now zoom in on the second part of **RQ2.2**: *how do search behaviours differ between individual group members?* From Table 4.6, a few of observations can be made about the individual search behaviours that follow our observations from the collaborative behaviour. As we can see on the right side of Table 4.6, the *prospectors* tend to dominate the query interactions during the search task in comparison with *miners* (*prospectors* issued 7.76 queries and *miners* 6.74 queries). Also for the querying behaviour, *peers* participants in **COL** condition participants issued more queries than *prospectors* and *miners* (on average 8.29 queries for *peers*) as well *individual* participants tend to issue more queries than *prospectors* and *miners* (on average 9.67 queries). Regarding interactions with documents, for the number of saved documents, we see that *prospectors* saved more documents than *miners* (from 10.83 saved documents for *prospectors* to 5.82 saved documents for *miners*, $p < 0.001$), for time spent reading documents, *miners* tend to spend more time reading documents than *peers* participants under **COL** condition (from 190 seconds for *miners* and 125 seconds for *peers*, $p = 0.021$). Finally, we observed that in the usage of the collaborative features, miners interacted more with the *recent queries* and *saved documents* components (from interface components ④ and ⑤ in Figure 2.3). From these observations, we can conclude that participants in the pre-defined assigned roles condition show compliance to what they were tasked to do. As outlined before, *miners* were tasked to spend more time curating the list of documents provided by the *prospectors*, and *prospectors* were tasked to find as many relevant documents as possible.

We now explore how artificial groups **INDM** and collaborative groups differ in terms of redundancy of querying behaviour and search space exploration. Joho et al. [69] showed that artificial groups had more redundancy of documents saved by individual participants than collaborative groups, and participants in artificial groups showed to have a less diverse search vocabulary. In Figure 4.6, we compute the overlap (more specifically the Jaccard Index) of the: *(a)* set of queries; *(b)* query terms; *(c)* clicked documents; and *(d)* clicked documents that are relevant according to the ROBUST05 relevance judgements. Here we explore clicked documents instead of *saved documents* as in SearchX's collaborative settings, a document can only be saved by one of the collaborators. In contrast to Joho et al. [69], artificial groups produced a lower overlap in terms of unique queries and consequently query terms. A possible explanation for that is that collaborative group members were influenced by each other's querying behaviour, which artificial groups were not. They tend to issue queries with similar query terms or even exact queries as seen on the *recent queries* widget. Regarding the search space exploration, we observe from Figure 4.6c-d that artificial groups have a higher overlap of clicks on the SERP than collaborative groups. We can conclude that awareness and sharing of knowledge about the exploration of the search space allow collaborative groups to be less redundant than artificial merged independent users when exploring the search space.

4.4.3 Communication Patterns

With **RQ2.3** we aim to explore to what extent the assignment of roles affects communication between group members. Table 4.7 provides an overview of the measures of communication patterns (as we defined them in Section 4.3.8) in our two collaborative

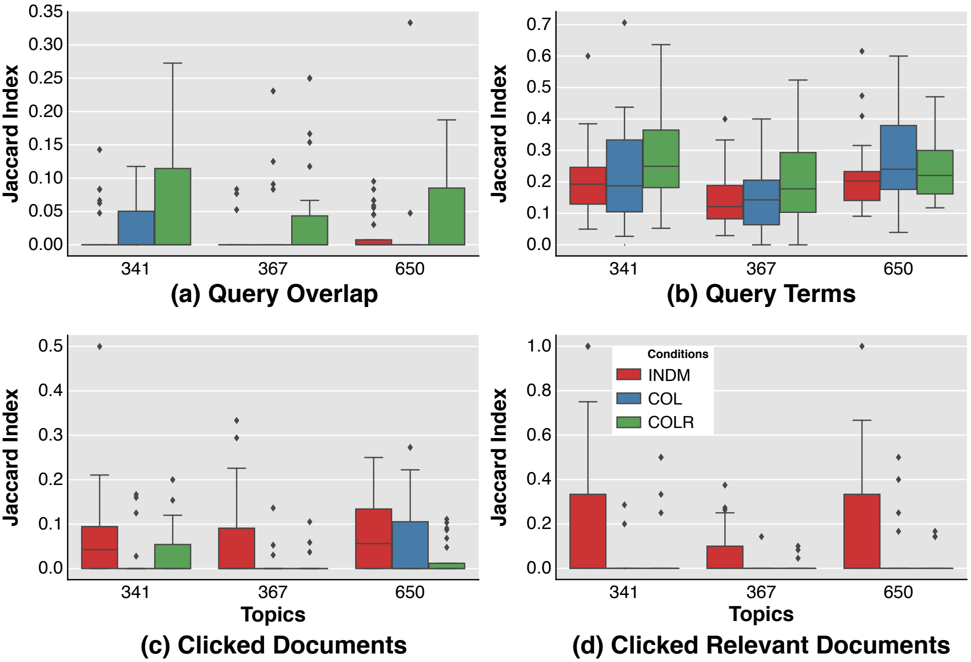


Figure 4.6: Overview of the Jaccard Index for the set of queries, query terms, clicks on documents, and clicks on documents that are relevant according to ROBUST05 relevance judgements between participants. Be aware of the different y-axis scaling.

conditions **COL** and **COLR**. From Table 4.7, in terms of number of messages exchanged among participants, in contrast to Tamine and Soulier [144], we found that participants in **COL** condition significantly send more messages than participants in in **COLR** condition. Also, participants without pre-defined assigned roles opened more the chat to communicate than participants with pre-defined roles, which are measured as *Effort* in Table 4.7. This was expected, as participants in **COL** were explicitly advised to use the chat tool to allocate their tasks. Consequently, groups in **COLR** make more use of the other collaborative widgets, as shown in Section 4.4.2.

To understand how the communication between group members was distributed, we show in Figure 4.7 the total number of messages over time. Most importantly we find that: (i) **COL** groups communicate much more than **COLR** groups; (ii) **COL** groups communicate throughout the entire three search topics in a consistently high volume; and (iii) **COLR** groups communicate mostly at the beginning of the first search topic. This can be explained by those participants in **COLR** only need to communicate for building an acknowledgement of each other in the beginning of the task.

We now explore the nature of the chat messages exchanged by group members. In Figure 4.8 we show the number of messages per group in the form of a boxplot, divided based on their coding (our coding scheme can be found in Table 4.4). We observe: that **COL** groups communicated much more about: (i) curating the list (code CL) of saved documents; and (ii) planning search strategies (code TC) than **COLR** groups. This is similar

Table 4.7: Overview of measures of communication patterns across the collaborative conditions. For each topic, the mean value is computed alongside standard deviation value. Statistical significant differences ($p < 0.01$) was found across the three topics for Volume, Effort, and Balance.

Topic	Cond.	Volume		Frequency		Effort		Balance	
341	**COLR**	5.00	(3.89)	29.69	(27.45)	13.50	(7.01)	0.72	(0.19)
	COL	10.29	(11.70)	45.58	(39.02)	15.36	(10.20)	0.58	(0.09)
367	**COLR**	4.00	(4.00)	64.15	(66.10)	9.77	(5.54)	0.76	(0.18)
	COL	9.19	(8.55)	43.71	(31.78)	16.25	(9.90)	0.59	(0.13)
650	**COLR**	5.50	(3.27)	77.66	(55.60)	12.40	(8.46)	0.75	(0.19)
	COL	14.18	(19.85)	61.25	(34.72)	17.59	(10.81)	0.62	(0.12)
All	**COLR**	4.74	(3.69)	59.61	(56.63)	11.58	(6.92)	0.75	(0.18)
	COL	11.32	(14.30)	50.61	(35.29)	16.47	(10.14)	0.60	(0.11)

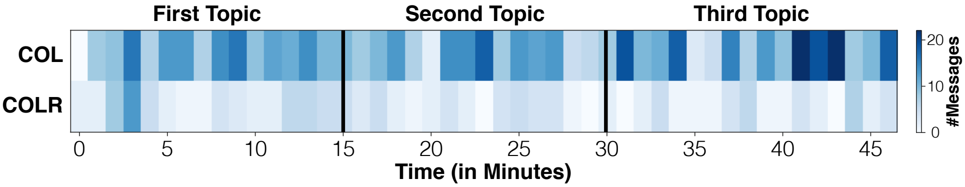


Figure 4.7: Timeline of the total chat volume across all **COL** and **COLR** groups respectively. For each one-minute interval, the number of messages in that interval is shown.

to the results found by Tamine and Soulier [144], as participants that are not assigned pre-defined roles will exchange more messages about the task and coordinate with each other.

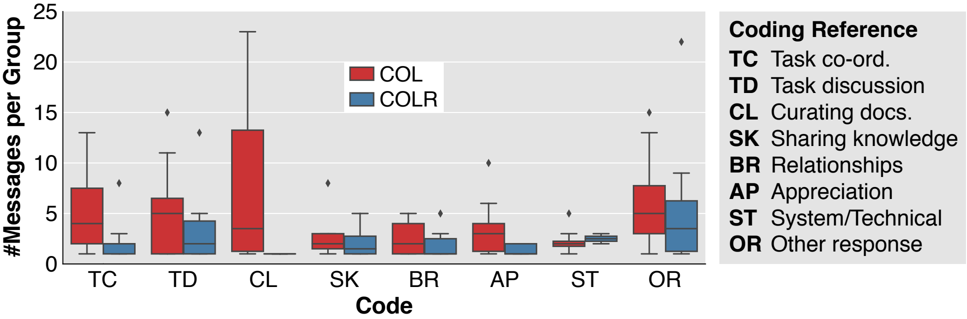


Figure 4.8: The number of messages exchanged for each of the eight classification codes, for each group, of the two collaborative conditions. A reminder of the coding references is provided to the right of the boxplot.

Overall, we thus find considerable differences in communication patterns between **COLR** and **COL** groups which are in line with our expectations for peer, prospector and miner communication patterns. However, despite this considerable difference in communication patterns, as we have seen, retrieval effectiveness is not exhibiting significant differences.

4.4.4 Participants' Previous Experiences and Perceived Effectiveness

Lastly, we analyse our participants' responses to the questions in the pre- and post-task questionnaires (the questions are listed in Sections 4.3.6 and 4.3.7).

In the pre-task questionnaire, we questioned our participants regarding their previous experience with collaborative search: 95 of our participants reported to have collaborated with other people to search the web before. From these participants, only three reported searching the web together with others daily, 17 weekly, 14 monthly, and 61 reported engaging in collaborative search episodes less frequently. The majority of the participants (66%) engaged in collaborative search in pairs, while the remaining engaged in groups of size three up to six members. We also asked participants about their last collaborative search experience. 36.8% of our participants reported to have collaboratively searched the web for travel planning (e.g., looking for hotels, cheap flight tickets, holiday cottages), 30.5% online shopping (e.g., comparing electronic products, looking for a house to rent or buy), 21% learning activities (e.g., literature researches, health searches), and the remaining participants collaborative work-related needs.

We now turn to the perceived effectiveness of our participants in the experiment. In the post-questionnaire, we collected participants' perceived search effectiveness, and collaborative search effectiveness. We also gathered feedback on the utility of each of the collaborative search feature described in Chapter 2. We summarise the participants' responses per experimental condition in Table 4.8.

Table 4.8: Summary of participants' responses to post-task questionnaire for questions using a Likert scale. For each experimental condition, the mean value is computed (with standard deviations).

	IND	**COL**	**COLR**
Overall system and task experience			
relevance of search results	3.56 (1.04)	3.65 (0.64)	3.69 (0.59)
system responsiveness	3.44 (1.45)	3.94 (0.78)	3.87 (0.88)
system technical issues	3.6 (1.32)	3.72 (1.02)	3.64 (1.07)
task difficulty	2.88 (0.73)	2.75 (0.71)	2.79 (0.8)
Collaborative search experience			
role assignment	–	4.13 (0.6)	4.42 (0.62)
communication with other members	–	3.75 (1.16)	3.46 (0.87)
awareness of each other's work	–	4.69 (0.44)	4.55 (0.53)
easiness of sharing knowledge	–	4.24 (0.74)	4.18 (0.83)
Collaborative search features feedback			
recent queries	3.0 (1.19)	4.31 (0.44)	4.22 (0.54)
saved documents	3.44 (1.19)	4.62 (0.45)	4.38 (0.58)
chat tool	–	3.85 (1.11)	3.09 (1.12)
hiding saved and excluded results	2.8 (1.19)	3.59 (0.75)	3.28 (0.9)

The first question in our post-task questionnaire is the perceived number of collaborators searching together, and it serves as a sanity check for our experiment. As we expected, the median response for this question was one collaborator for the collaborative conditions **COLR** and **COL** (min. 0 and max 2), and for the independent condition **IND**, the median response was zero collaborators (min. 0 and max 1). The next four questions in our post-questionnaire focus on the participants' overall system and task experience. The feedback for these questions is similar across the experimental conditions. Regarding the overall system experience, feedback ranged from 3–4 on the 5-point Likert scale, and thus we can conclude that the overall search experience was positive. Similar to Tamine and Soulier [144], for the task difficulty question, we also find similar figures across the collaborative experimental conditions and a slight increase in difficulty for the **IND** experimental condition. Concerning the perceived experience of the collaboration experience questions, we can also observe similar values for the collaborative experimental conditions. This is in contrast with Tamine and Soulier [144] in which they found that following the roles instructions was the most difficult for groups with pre-defined assigned roles; however, we did not ask participants an open-question regarding their difficulty following the task instructions. Finally, concerning the collaborative search features feedback, participants agreed that the saved documents were the most useful feature for all the collaborative conditions, followed by the recent queries. The chat tool was considered the least useful feature by the **COLR** participants in contrast with those in **COL**. This observation can be explained by the results shown in Table 4.7: participants in the **COLR** condition communicate less and make less use of the chat tool.

4.5 Reflections

4.5.1 Search Effectiveness

In this chapter, we employed precision and recall as our target measures of search effectiveness. As our analyses have shown, we did not observe a significant difference between groups with and without pre-defined assigned roles in terms of those measures. This means that our results have different takeaways from prior works, and we could not confirm the same findings by Pickens et al. [111] and Tamine and Soulier [144]. In particular, we could not find improvements of **COLR** over **INDM** (see row 2, Table 4.1) as found by Pickens et al. [111] (see row 2, Table 4.1). Moreover, we could not find improvements in terms of precision of **COL** over **COLR** (see row 2, Table 4.1) as found by Tamine and Soulier [144]. The following are differences in our study that could explain the different results in terms of search effectiveness found in this chapter:

- We conducted our work in a crowd-sourcing environment, where participants in a group did not have to know each other before the experiment. We did not require participants to know each other as it is a work-task designed to be conducted independently of previous inter-personal experience among participants.

- In contrast to Tamine and Soulier [144], we assigned the same three search topics to participants in all the conditions. In Tamine and Soulier [144], each condition had a different search topic.

- In contrast to Tamine and Soulier [144] but similar to Pickens et al. [111], we did not let participants with pre-defined assigned roles choose their role preference.

4.5.2 Search Behaviour and Communication Patterns

We explored the search behaviour of collaborative search users with and without assigned roles. In Table 4.6, we analysed a variety of individual and collaborative search behaviour. Despite differences in behaviour between groups with assigned roles and groups without pre-defined roles, we did not see a change in retrieval effectiveness. In Figure 4.6, our results also suggest that awareness and sharing of knowledge features (e.g., recent queries, and saved documents) allow collaborative groups to be less redundant than artificial merged independent users when exploring the search space.

In terms of communication patterns, we found considerable differences in communication patterns between **COLR** and **COL** groups, which are in line with our expectations for peer, prospector, and miner communication patterns. However, despite these considerable differences in communication patterns, retrieval effectiveness did not exhibit significant differences.

4.5.3 Implications for Design

Considering our findings, we find that there is room for designing user studies that investigate the composition and interaction of these two roles for groups above size two as was conducted in the previous chapter. For instance, it remains an open question how adding a *prospector* or *miner* to a group would affect search effectiveness—one can hypothesise that adding a *prospector* may improve recall while adding a *miner* may improve precision.

According to our search behaviour and communication pattern analysis, there is an opportunity for designing systems that are able to predict the task intent during the search session. For instance, we observe that the number of curating messages tends to be more frequently towards the end of the search session. Hence, one line of work is to design collaborative features that could support the automatic splitting of the saved documents list. Therefore, such a feature could better drive the collaborative search task and make it more structured.

Although pre-defined assigned roles may constrain the participants considerably, they are required for some collaborative search tasks, especially work-tasks. From our findings, we observed that assigning user roles reduced communication, but it did not translate into improvements in search effectiveness.

In our study design, we included in our task description the roles of the collaborators (e.g., for the *prospector* we shortly described the goals of their *miner* collaborator). One future step in this line of work is to design interfaces that could bring collaborators' awareness of each others' roles, for instance, automatically supporting users' roles given their explicit preference to one of the roles during the search session.

4.6 Limitations

In this section, we analyse the limitations of this chapter as our experimental setup had limitations related to: (i) the effect of the random pre-defined role assignment; (ii) the number of participants in each experimental condition; (iii) user study instructions; and (iv) crowd-sourcing experiments in times of a global crisis. First, the random pre-defined role assignment could have affected the groups with pre-defined roles. We acknowledge that a participant could have a preference for one of the assigned roles, which could have affected the overall search performance of a collaborative group. Our results could have been affected by the choice of these roles we chose in this work. Also, a participant could be skilled in one of the assigned roles. Second, the number of participants (120) was determined mainly by considering the total cost of the study (£1350.00). We acknowledge that the number of participants could be limited, but we believe that it does not affect the importance of the presented results, but leaves space for future work. Third, we designed our user study instructions based on previous work in collaborative search and guidelines in HCI for information retrieval [72, 128, 144]. We acknowledge that the instructions were given to our participants as a limitation to this study. Given that the instructions defined the roles that we selected for this work, we observed that the participants followed our guidelines as we expected. Lastly, we recruited participants for this experiment in a period of the Coronavirus pandemic (Europe lockdown in April 2020). We acknowledge that the increased distractions at home and workplace (e.g., with homeschooling or family care) could have affected the participants' performance during the experiment since most of our participants were located in countries with rigid lockdown measures (most of our participants resided in the UK).

4.7 Conclusions

In this chapter, we have explored a group configuration of CSE, the use of pre-assigned *roles* in collaborative search. Previous works have provided opposing evidence as to the utility of assigned roles in collaborative search. The main weakness of these studies was that each one was limited to a comparison of two experimental conditions. We set out to settle the question of the utility of role-based collaborative search. In particular, we aimed to answer the following research questions:

RQ2.1 Does retrieval effectiveness between collaborative groups with unassigned and pre-assigned roles differ?

RQ2.2 Given the collaborative groups outlined in **RQ2.1**, how do search behaviours differ between them, and compared to searchers working individually?

RQ2.3 To what extent does the assignment of roles affect communication between group members?

We designed a crowd-sourcing based user study across four conditions: (i) individual searchers; (ii) artificial groups formed by aggregating the search logs of pairs of individual searchers; (iii) pairs of collaborating searchers without assigned roles; and (iv) pairs of collaborating searchers with assigned roles. Aside from search effectiveness, we investigated the differences in search behaviour and communication patterns between those conditions.

We found that search effectiveness does not differ significantly between groups with and without assigned roles, even though our search behaviour and communication pattern analyses exhibit a substantial difference among those groups. Specifically, we found that our retrieval effectiveness measures (precision and recall) do not differ significantly between groups with and without assigned roles. Artificial groups (i.e., pairs of individual searchers) who do not benefit from communication and coordination perform at the same level of recall as collaborative groups. We found that groups with assigned roles show search behaviours that indicate their compliance with the given role assignment: group members tasked with exploring the search space spent less time reading documents, and issued more queries. Differently, group members tasked with exploring the set of documents pre-selected by another group member spent more time reading documents. Despite the apparent differences in search behaviours, search performance overall does not differ significantly, as indicated above. Lastly, we found that groups with assigned roles communicate less frequently than groups without assigned roles. Still, the time saved from fewer exchanges does not turn into an improved search performance. We conclude that role-based collaborative search—at least when employing the roles as established in the literature—is not advantageous for collaborative search. Users are similarly successful when self-organising the collaborating group.

Future directions to extend our work should include search tasks that take longer periods of time and potentially several search sessions to complete in collaboration—do the results hold when the search tasks become significantly more complex? In addition, another line of work is to examine conversation agents that aids collaborators in the organisation and alignment of their search strategies and their division of labour (research initiated by [9]). Lastly, we have so far considered two roles (prospector and miner) that are well established in the collaborative search literature [53]. A limited number of other types of roles have been proposed in the past that have been shown to be effective in *simulated* settings (e.g., domain expert and domain novice [136])—a future work here is to translate those simulations into user studies to determine the practical utility of those role types.

In this chapter and Chapter 3, we have analysed CSE systems group configurations. In the following chapters, we look at CSE with a different perspective, where CSE systems are used in the contexts of search as learning and collaborative shopping.

5

Contrasting Search as a Learning Activity with Instructor-designed Learning

In previous chapters, we focused on the group configurations such as group size *in Chapter 3 and* group member's role *in Chapter 4. In this chapter, we bring into focus the usefulness of collaborative search surrounding* human learning *during the search process. In particular, we are interested in answering whether a (single-user or or collaborative) search session is as effective as a lecture video—our instructor-designed learning artefact—for learning. To answer this question, we designed a user study that pits instructor-designed learning (a short high-quality video lecture as commonly found in online learning platforms) against three instances of search, specifically (i) single-user search, (ii) search as a support tool for instructor-designed learning, and, (iii) collaborative search. The contributions of this chapter have been published in [95].*

5.1 Introduction

Search as Learning is a research area within information retrieval that considers questions surrounding *human* learning during the search process: how much or how little do users learn while they search and in what ways can search technology be adapted and optimised for human learning? In his seminal paper, Marchionini [88] remarked on the importance and complexity of what he called *learning searches* (i.e. search activities for the purpose of human learning), a subset of exploratory search: "*Learning searches involve multiple iterations and return sets of objects that require cognitive processing and interpretation. These objects [...] often require the information seeker to spend time scanning/viewing, comparing, and making qualitative judgements.*" At the Second Strategic Workshop on Information Retrieval in 2012 [1] search as learning was recognised as an important future research direction that will help "*people achieve higher levels of learning through [...] more sophisticated, integrative and diverse search environments*". This call for research has been taken up in recent years by a number of researchers in several directions, including optimising retrieval algorithms for human learning [28, 34, 82, 107, 121, 141, 142], observing how users currently make use of search engines for learning-oriented information needs [25, 45], developing metrics to measure the amount of learning taking place during the search process [162], and arguing for more reflective search behaviour ("slow search") in contrast to the current demands for an instant—and increasingly proactive—search experience [148].

Search and sensemaking is an intricate part of the learning process, and for many learners today synonymous with accessing and ingesting information through Web search engines [22, 106, 146]. At the same time, Web search engines are not built to support users in the type of complex searches often required in learning situations [54, 59, 88]. But what effect does this lack of a learning-focused Web search engine design have on the ability of users to learn compared to a setting where they are provided with high-quality learning materials? In this chapter we set out to answer this question by *measuring* how effective searching to learn is compared to (i) learning from—in our experiment: high-quality video—materials specifically designed for the purpose of learning, (ii) learning from video materials in combination with search, and, (iii) searching together with a partner to learn (i.e., CSE for learning).

The aim of our work is to *quantify* to what extent search as a learning activity is a viable alternative to what we call *instructor-designed learning*, that is, learning materials designed and created specifically for the purpose of learning. As not for every possible topic specifically designed learning materials exist, it is important to understand what effect that has on one's ability to learn. In addition, we are also interested in understanding whether the lack of learning materials can be compensated in the search setting by the presence of a second learner that has the same learning intent (i.e. CSE for learning).

Our work is guided by the following research questions:

RQ3.1 How effective (with respect to learning outcome) is searching to learn compared to instructor-designed learning?

RQ3.2 How effective (with respect to learning outcome) is instructor-designed learning supported by search in comparison to just instructor-designed learning?

RQ3.3 How effective is pair-wise CSE compared to single-user search for learning?

Specifically, in this work we conducted a user study with 151 participants and measured *vocabulary learning*, a particular instance of human learning (similar in spirit to [141, 142]), across five search and instructor-designed learning conditions. As high-quality instructor-designed learning materials we make use of lecture videos sourced from TED-Ed, Khan Academy and edX, popular online learning platforms. Our main findings can be summarised as follows:

- We find participants in the instructor-designed learning condition (watching high-quality lecture videos) to have 24% higher learning gains than participants in the searching to learn condition;
- CSE as learning does not result in increased learning gains;
- The *combination* of instructor-designed learning and searching to learn leads to significantly higher learning gains (an increase of up to 41%) than the instructor-designed learning condition without a subsequent search phase.

5.2 Related Work

We now provide an overview of the areas related to our work: exploratory search, search with an educational intent and collaborative search.

5.2.1 Exploratory Search

Exploratory search tasks are often complex, open-ended and multi-faceted [159]. They tend to span several sessions and require next to finding, the analysis and evaluation of the retrieved information. Marchionini's overview of exploratory search challenges and opportunities [88] marked the beginning of a long series of related workshops and evaluation campaigns that continue to this day [18, 158–160]. Several works have characterised users' search behaviours in this setting. Athukorala et al. [5] investigated to what extent simple lookup tasks differ from exploratory search tasks with respect to easily measurable behaviours such as the initial query length, the time spent on analysing the first SERP, the scroll depth and task completion time. Later, Athukorala et al. [6] leveraged their positive findings (these tasks do indeed differ in several behaviours) and proposed a robust predictor that determines based on the first traces of a search session whether the session will end up being of an exploratory nature.

Besides analysing users' exploratory search behaviours, a number of studies have focused on developing user interfaces and algorithms to support complex information needs, e.g., [54, 59, 122]. Golovchinsky et al. [54] proposed several interface elements to better support multi-session search, with a heavy focus on visualising the query history and query patterns, while Ruotsalo et al. [122] presented an interactive intent modelling interface to simplify the process of moving the exploration into one direction or another. On the algorithmic side, Hassan Awadallah et al. [59] explored an automated approach (based on query logs) towards decomposing complex search tasks into relevant subtasks, a step of the search process that, in current Web search engines, is largely left to the user.

Our work is in line with prior search behaviour observation studies: we create different learning conditions and then observe and analyse our participants' behaviours in a relatively common Web search setup. One particular type of exploratory search are learning searches [88], which in recent years have been explored under the search as learning heading [37] as we discuss next.

5.2.2 Search as Learning

Information scientists have observed that learners of all ages increasingly turn to search engines to support their learning [50, 106, 118]. At the same time, concerns have been raised about the lack of individuals' "*critical and analytical skills to assess the information they find on the Web* [118]."

Several works have explored data-driven methodologies to determine the impact of (developing) expertise on search behaviour [45, 161] and subsequently to exploit measurable behavioural traces (log traces, eye-tracking traces) as proxies of domain knowledge [35, 169]. Relying on users' log traces and features derived from them (e.g., query complexity, diversity of domains on the SERP, document display time) enables the use of a large user population (e.g., more than 700K search sessions in [45]); at the same time though, these heuristics can only be considered to be crude proxies of learning gain metrics (i.e. the difference between the knowledge at the end and the start of the search session) and they require large-scale log traces to overcome the variance of the user population. Instead of relying on search behaviour proxies, some works have measured learning directly through the explicit assessment (e.g., through multiple-choice tests, mind maps, the writing of a summary) of domain knowledge before and after the search as learning session [36, 39, 141, 162]—this of course is only viable in a lab setting with a limited set of users. In this chapter, we follow the latter line of prior works, conducting a user study and measuring learning gains by assessing our participants' before and after the learning session.

The main setup of our study is inspired by [36, 141, 142]. Collins-Thompson et al. [36] conducted a user study to investigate whether certain search strategies (single-query, multi-query, and intrinsic-diversified search results) are conducive to learning. They measured learning outcomes via manually assessed open-ended questions as well as self-reports and found both to correlate highly. Syed and Collins-Thompson [141] introduced a document ranking model optimised for learning (instead of relevance as standard ranking models) and showed it to be more beneficial than standard retrieval algorithms with respect to learning outcomes. This finding though is based on a rather artificial study setup: the study participants were provided with a fixed list of ranked documents (produced by variants of the document ranker) on a given topic that they were required to read, before answering knowledge assessment questions—the user study explicitly avoided the use of an actual search engine and the associated typical search behaviour (issuing several queries before clicking a document, skipping over documents in the ranked list, etc.). In the work we present here, we investigate a more realistic setup, with topics drawn from online learning platforms and search sessions that require our participants to search the Web as they would usually do. Importantly, we compare the effectiveness of learning not just within search variants but also with respect to instructor-designed learning material.

5.2.3 Collaborative Search

In addition to single-user search variants, we also explore CSE (i.e. multiple users collaborating in the search process) in our study. The inclusion of this variant stems from the fact that collaborative searches for highly complex information needs, as may be encountered during learning, can yield significantly better results with respect to material coverage and knowledge gain when conducted in collaboration [101, 129, 134].

A number of CSE systems have been proposed in the past [30, 54, 68, 84, 102, 109, 128, 163, 167], though few of those systems are still accessible and functioning today. They all have been designed with a number of goals in mind, the most essential ones being (i) *awareness* of each others' actions (e.g., through a shared query history), (ii) enabling the *division of labour* (e.g., through algorithmic approaches [129] or a chat to explicitly divide the work), and (iii) *knowledge sharing* so that the collaborators do not duplicate their work (e.g., through shared bookmarks).

Lastly we note that systems can support different types of collaborations. As discussed in Chapter 2, there are four main dimensions of CSE episodes: *intent* (explicit or implicit collaboration), *depth* (algorithmic changes to support CSE vs. user interface changes), *concurrency* (synchronous vs. asynchronous) and *location* (remote vs. co-location). In our work, we designed our CSE system to be used in an explicit collaboration, with changes restricted to the user interface level and remote users collaborating in a synchronous manner. These choices are not only governed by our user study setup, but also the fact that those are the most common characteristics of existing CSE systems.

5.3 Experimental Design

We set up our study as a *vocabulary learning task* which requires study participants to recall and produce the meaning of domain-specific terms. This task enables us to measure the *learning gain*—the dependent variable in our study—effectively and efficiently as the difference between the vocabulary knowledge in a pre- and post-test. Importantly, this task can be executed within a short time frame—such as a single search session—permitting us to recruit crowdworkers for our study as also previously done in [141, 142]. Learning tasks with more cognitively complex activities such as *create* or *design*, in contrast, require longitudinal studies (e.g., [25]) and considerable more assessment efforts to judge the artefacts created during learning (e.g., summaries [162]).

We now describe the SearchX Setup we used in our study, how we selected the topics for our study, then discuss metrics to measure vocabulary learning and finally present the five different experimental conditions we evaluated in our work.

5.3.1 SearchX Setup

We used our search system SearchX(see Chapter 2, version V1) with the back-end connecting to the Microsoft Bing API to serve high-quality Web search results (see Figure 2.3). For single-user search participants, we provided SearchX with a search box with four search verticals (Web, Image, Video, and News) ①, a document viewer (see Figure 2.8) without document rating ⑨ and document annotation ⑩, recent queries ④, and saved documents ⑤. For the synchronous CSE of two or more users, we also included the interface elements recent queries ④ and saved documents ⑤ but with colour coded for each group member (see Figure 2.6a) and group chat ⑥. [1]

5.3.2 Search as Learning Topics

One particular setting where we envision search as learning to play an important role is online learning—video lectures are widespread today and a vital component of the increasingly popular Massive Open Online Courses (MOOCs). Choosing high-quality lecture videos on very specific topics that were designed (often by instructional designers in the case of MOOCs) for learning makes the search challenge hard—enabling us to get a realistic answer to our research questions. Initially we chose three large-scale sources of lecture video content: Khan Academy[2], edX[3] and TED-Ed[4]. From both TED-Ed and Khan Academy we selected ten of the most popular videos (more than half a million views each); on the edX platform we first selected ten of TU Delft's STEM MOOCs at undergraduate level and then selected a lecture video from within the first two course weeks that was no longer than 15 minutes. The selected candidate videos cover a range of topics including *dystopia*, *stoicism*, *magnetism*, *photosynthesis*, *radioactive decay* and *climate change*. We manually created a vocabulary list for each of the in total thirty selected videos—a term entered the vocabulary list if (i) it was mentioned in the video at least once and (ii) it does not frequently occur outside of the domain-specific context as judged by the two annotators. This resulted in vocabulary lists with a median size of 30 items (minimum 23, maximum 73).

As such large lists were not feasible to be used in our actual study, we filtered the videos and vocabulary items by their difficulty and only retained the ten videos and their respective ten *most difficult* vocabulary items. Here, we employed the *amount of unfamiliar terminology* in a video as a proxy of video difficulty. In order to ascertain the difficulty of the videos and vocabulary respectively we asked three staff members of our institute (all with a PhD in computer science) to label all of the vocabulary items with a score between 1 (akin to *unknown term*) and 4 (akin to *I know the meaning*)[5]. The labellers only received the vocabulary list, not the corresponding video. This resulted in a *vocabulary knowledge score* per video, which is simply the average score of all vocabulary items across the three labellers. We then ranked the videos according to their average score and selected the ten videos with the *lowest* scores, i.e. those with the largest amount of unknown terminology according to our labellers. Similarly, we also ranked each video's vocabulary items according to the average score across the three labellers and retained the ten least known ones. The final list of videos (identified by their topic), as well as a selection of the retained vocabulary items are listed in Table 5.2. The vocabulary items shown are highly domain-specific, a setup that contrasts with [141, 142] where participants' vocabulary knowledge was also tested on less domains-specific vocabulary such as "temperature" and "earth". The majority of videos in Table 5.2 are from edX; the average video length is 7.3 minutes, a common length of MOOC lecture videos [56].

[1] Here we used the group chat instantiated with `Converse.js` (see Figure 2.5a).

[2] https://www.khanacademy.org/

[3] https://www.edx.org/

[4] https://ed.ted.com/

[5] Concretely, we employed the Vocabulary Knowledge Scale as outlined in Section 5.3.3, but did not require our labellers to actually write down the meaning of the items identified as knowledge levels (3) or (4) due to the sheer size of the vocabulary list.

This topic/video selection process ensures that our study participants are likely to find at least one of our topics unfamiliar with a high potential for vocabulary learning. That this is indeed the case, is visible in Figure 5.3—on average more than half of the tested vocabulary terms were unknown (knowledge levels 1 or 2) to the study participants.

5.3.3 Assessing Vocabulary Knowledge

We employ the *Vocabulary Knowledge Scale* (VKS) test [43, 157] as it has been shown to be a reliable indicator of vocabulary knowledge. The VKS tests the incremental stages of word learning [40] with the following statements:

1. *I don't remember having seen this term/phrase before.*
2. *I have seen this term/phrase before, but I don't think I know what it means.*
3. *I have seen this term/phrase before, and I think it means ___.*
4. *I know this term/phrase. It means ___.*

We employ statements (1) to (4) to test the vocabulary knowledge for each of our vocabulary items[6]; the latter two statements require our study participants to recall and reproduce the meaning of the vocabulary item. Choosing statement (3) indicates uncertainty about the meaning's accuracy, statement (4) indicates certainty on the correctness of the provided meaning. In order to investigate to what extent this self-assessment is correct among the crowdworkers that participated in our study (Section 5.3.6 provides more information on them), we randomly sampled 100 of the meanings written by our participants across all vocabulary items—fifty from participants self-reporting levels (3) and (4) respectively. We manually labelled the statements as either *incorrect*[7], *partially correct*[8] or *correct*[9]. The results in Table 5.1 show that 88% of the statements self-assessed at knowledge level (4) are either correct or somewhat correct. At level (3), this holds for 68%. These results indicate that the self-assessment scores are robust and thus we use them without further manual labelling of the more than 3000 assessed vocabulary items. This is in line with the study conducted in [36], where users' perceived learning outcomes (i.e. the self-assessment) matched closely the actual learning outcomes (i.e. the produced definitions). Finally, it is worth pointing out that this setup is more difficult to tackle (as it requires the *production* of a definition) than closed multiple-choice questions (which require the *recognition* of a definition) as employed to test vocabulary learning in prior work [141, 142].

5.3.4 Learning Metrics

As [142], we report *absolute learning gain* (*ALG*) and *realised potential learning* (*RPL*), enabling us to directly compare our study results to prior works. *ALG* is the aggregated

[6] Note, that the VKS test also contains a fifth statement geared towards second language learners. As in our study we only include native English speakers, we ignore it here.

[7] Incorrect example: *superposition* (Qubit topic) described as "this has to do with the linear system".

[8] Partially correct example: *Bra* (Qubit topic) described as "vector".

[9] Correct example: *propofol* (Anesthesia topic) described as "an inhalation anesthetic used to induce sleepiness".

Table 5.1: Labelling of 100 sampled VKS level 3/4 statements.

	Correct	Partially Correct	Incorrect
VKS level 3	42%	26%	32%
VKS level 4	76%	12%	12%

difference in knowledge observed in the post- and pre-test across all vocabulary items $v_1,...v_m$. Here, $vks^X(v_i)$ is the knowledge score assigned to v_i; X is the test (pre or post). As knowledge state changes from level (1) to (2) between pre- and post-test are natural (after the pre-test, each item has been seen at least once), we collapse the two lowest levels and assign both a score of 0. Items at knowledge levels (3) and (4) are treated in two ways: (i) in the binary setup we treat items at both levels in the same manner and assign a score of one; (ii) in the more fine-grained setup we assign scores of 1 and 2 respectively. The advantage of the binary setup is a more intuitive explanation of the ALG/RPL metrics as we will see later. We also assume that knowledge does not degrade between the pre- and post-test. ALG is then computed as follows:

$$ALG = \frac{1}{m}\sum_{i=1}^{m} max\left(0, vks^{post}(v_i) - vks^{pre}(v_i)\right) \tag{5.1}$$

The RPL metric normalises ALG by the maximum possible learning gain (MLG) of each item (either 1 in the binary case or 2 in the fine-grained setup):

$$MLG = \frac{1}{m}\sum_{i=1}^{m} maxScore - vks^{pre}(v_i) \tag{5.2}$$

$$RPL = \begin{cases} \frac{ALG}{MLG}, & \text{if } MLG > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.3}$$

To compute the metric for a particular condition, we average the metric across all participants in that condition. We determine statistical significance through the non-parametric Kruskal-Wallis rank test which allows for the comparison of more than two groups.

5.3.5 Experimental Conditions

Figure 5.1 provides an overview of the study design we employed. Across all conditions, every participant first conducts a *pre-test* for which randomly three of our ten final topics are selected; for each of those topics all ten vocabulary items are assessed as described in §5.3.3. The participant is then assigned the topic $T_{difficult}$ for which she reported the lowest average knowledge levels. In case of a tie between topics, we randomly pick one. After that, the participant is randomly assigned to one of the conditions. The experiment ends with the post-test, in which the participant is again assessed on her vocabulary knowledge—this time only for items of $T_{difficult}$. In addition, the post-test also requires the participant

Table 5.2: Overview of topics/conditions. Conditions: [SE] search only; [V] video only; [V+SE20] video followed by 20 minutes of search; [V+SE] video and search totalling at least 20 minutes; [CSE] collaborative search. The three right-most columns contain three examples (items at difficulty rank 1, 5 and 10) of the ten vocabulary items within a video.

	Participants per Condition						Video	Avg.	Vocabulary item difficulty rank		
Topic	SE	V	V+SE	V+SE20	CSE	Source	length	VKS	1 [most difficult]	5	10 [least difficult]
Radioactive decay	4	5	7	5	10	edX	6m53s	2.72	*Auger electron*	*K-shell electron*	*electron capture decay*
Qubit	5	3	5	2	2	edX	12m24s	2.81	*Ket*	*superposition*	*quantum information*
Water quality aspects	2	6	2	4	0	edX	10m45s	2.88	*trihalomethanes*	*bacteriophages*	*blue baby syndrome*
Religions	0	1	1	0	0	TEDEd	11m09s	2.91	*dharma*	*compendium*	*pilgrimage*
Sedimentary rocks	3	0	0	2	6	edX	5m03s	2.92	*feldspars*	*mud flats*	*sedimentary rocks*
Anesthesia	4	3	6	2	2	TEDEd	4m55s	2.94	*sevoflurane*	*diethyl ether*	*opium poppy*
Glycolysis	5	4	3	6	20	Khan	13m29s	2.97	*krebs cycle*	*electron transport chain*	*cellular respiration*
Urban water cycle	1	1	0	0	2	edX	7m40s	3.01	*Lesoto Highlands*	*coagulation*	*recontamination*
Depression	0	0	1	0	0	TEDEd	4m28s	3.02	*norepinephrine*	*transcranial*	*cholesterol*
Industrial biotech	2	2	0	4	8	edX	5m48s	3.02	*tobacco mosaic virus*	*prokaryotic*	*fungi*
#Participants total	26	25	25	25	50						

to write a short summary on the topic as well as an outline[10]. In the CSE experiment, the two collaborators independently perform the pre-test and the post-test and collaborate during the CSE phase. For collaborating users we slightly extended the pre-test phase: we provided examples of collaborative searches and added seven questions on their past collaborative Web search experiences—taken from a large survey on collaborative Web search [101]—in order to reinforce the collaborative nature of the upcoming task.

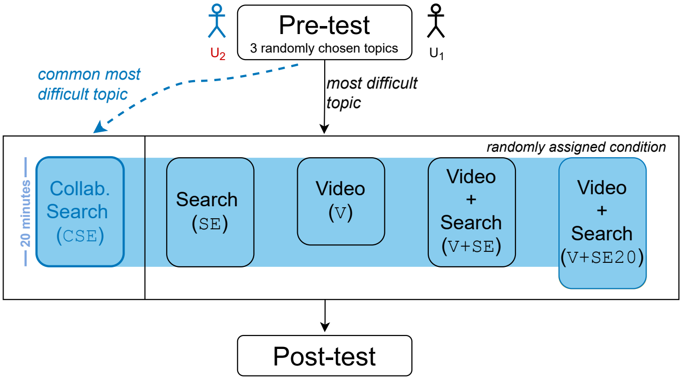


Figure 5.1: Study design overview: four single-user conditions and one collaborative (pairwise) condition.

Our participants are randomly assigned to one of five conditions:

Video (V) In this condition, a participant is given access to the lecture video and can watch it at her own pace (the common video player functions pause, rewind and skip are enabled).

Search (SE) Here, the participant is provided with the single-user search interface and instructed to search on the assigned topic *for at least 20 minutes*.

Video+Search (V+SE) The participant first views the video as in the **V** condition and afterwards is provided with the single-user search interface and asked to search on the assigned topic. The minimum time for this task (across both video watching and searching) is 20 minutes.

Video+Search (V+SE20) This condition is similar to **V+SE**, the only difference is that now the participant is instructed to spend 20 minutes searching after having viewed the video.

Collaborative Search (CSE) Two participants search together using the collaborative version of our search system for at least 20 minutes. Despite in Chapter 3 we have varied group sizes, here we used group size two as it is a typical setup in online collaborative learning in preliminary studies [77].

For all conditions involving a search phase, we employed the task template in Figure 5.2, adapted to our use case from previous studies [36, 79].

[10] While collected, we leave the analyses of the outline and summary to future work.

> Imagine you are taking an introductory **[general topic, e.g., *Health and Medicine*]** course this term. For your term paper, you have decided to write about **[specific topic covered in the video e.g., *the symptoms and treatments of depression*]**.
> The professor requires all students to watch a course video about a particular topic. Then, the students have to demonstrate what they learn about a particular topic by collaboratively conducting searches online and presenting their views on the topic. To prepare your term paper, you and your partner need to collect and save all the webpages, publications, and other online sources that are helpful for you to write a paper. After you and your partner have completed watching the course video and the search phase, you will be asked to complete 13 exercises; those exercises include questions about your term paper topic and the writing of an outline for your term paper. Those exercises are solved individually (without your partner).

Figure 5.2: Task template for all conditions containing a search phase. The underlined green phrases were only added in the **CSE** condition; shown in dashed orange are the instructions only added for the **V+SE** and **V+SE20** conditions.

In the video-only condition (**V**) we instruct our participants to watch the video without any mention of search. Note, that the task description above does not explicitly state the nature of the post-test (ten of the thirteen "exercises" are our vocabulary learning questions), instead the focus is on acquiring on overview of the specific topic.

Apart from the video-only condition, all other conditions have a minimum task time; the participants are provided with a visible timer, and can complete the post-test as soon as the required time on the task is reached. We settled on a twenty minute task time to provide participants with sufficient time to search and learn while keeping the study time feasible for crowdworkers. We added three compliance steps in our study design: (i) we included a *sports* topic (with well-known vocabulary items such as *football*, *winner*, etc.) in the pre-test and excluded workers who chose knowledge levels 1/2 here; (ii) we disabled copy & paste and recorded all tab changes in the pre- and post-tests and alerted participants to the fact that more than three tab changes lead to non-payment (to avoid participants searching the Web for answers to the questions); we limited the tab changes in the video watching period to three changes as well; and (iii) we required participants to adhere to a minimum word count in the open questions of the post-test.

We arrived at this design after a number of small pilot studies on the crowdsourcing platform FigureEight (formerly CrowdFlower). As FigureEight is mostly suitable for short tasks, we performed the actual experiments on the Prolific Academic platform, which has been shown to be a more reliable source of workers for cognitively demanding tasks than FigureEight [110].

5.3.6 Study Participants

Over the course of 27 days, in January 2018, a total of 151 study participants completed our experiment successfully across the five conditions on Prolific. Their median age was 31 (minimum: 18, maximum 66). 62.7% of our participants were female; most participants are from the UK (70.9%), the remaining participants are from Australia, the USA and Canada. Their academic backgrounds varied: 43.0% reported a high-school diploma as highest academic degree, 35% an undergraduate degree and the remaining 22% a graduate degree. We paid our study participants £5.00 per hour for the experiment, which was the price suggested by the Prolific platform. The median time they spent in our experiment (including the pre- and post-tests) was 49 minutes.

As the **CSE** condition is set up as a synchronous CSE task (i.e. two study participants have to be online at the same time), we added a waiting period for at most 10 minutes at the end of the pre-test; if within that time, no other participant completed the pre-test with the same topics, we released the participant from the task and paid £1.25 for a completion of the pre-test and the waiting period only.

Finally, we note that next to the 151 valid ones, we rejected 20 submissions—these participants did not adhere to our compliance standards (such as at most three tab changes). We continued the crowdsourcing task until we reached at least 25 participants/pairs for each condition. The relatively low number of rejections despite the complexity and length of the task indicates Prolific to be a suitable platform for this type of user study.

5.4 Results

We now discuss the results, organised according to the three research questions.

5.4.1 Search vs. Instructor-Designed Learning

In **RQ3.1** we investigate whether search as learning is as *effective* as instructor designed learning, that is, as effective as watching the lecture video. We thus focus on comparing conditions **SE** and **V**. We computed the absolute learning gain and realised potential learning for each study participant in the binary setup[11]; the results, averaged across participants of a condition, are shown in Table 5.3. In the **V** condition, the average *ALG* is 0.32, that is, on average the participants increased their knowledge on three out of ten vocabulary items from knowledge levels 1/2 to levels 3/4. The interpretation of *RPL* is equally intuitive: for the **SE** condition for example, this metric is 0.3, indicating that on average the participants reached knowledge levels of 3/4 for thirty percent of the terms that were unknown to them.

When comparing **SE** with **V**, although we do not observe a statistically significant difference between the two (recall that the sample size overall is not very large), the results show a trend: instructor-designed learning leads to a 14% (measured in *ALG*) and a 24% (*RPL*) increase in learning gains respectively. Practically, the change in *RPL* from 0.3 to 0.37 means that participants in the **V** condition reached knowledge levels 3/4 for "almost" one more vocabulary item than participants in **SE**.

In Figure 5.3 we zoom in on the knowledge state changes between the pre- and post-test and report the fraction of the most important types of changes. Across all conditions, participants in the **SE** condition have the largest percentage (40.38%) of vocabulary items that remain at knowledge levels 1/2 in the post-test. As expected, in the **V** and the two

[11] We found the same statistical differences in the more fine-grained setup (which distinguishes knowledge levels 3 and 4); here we report the binary case as it is more intuitive to interpret.

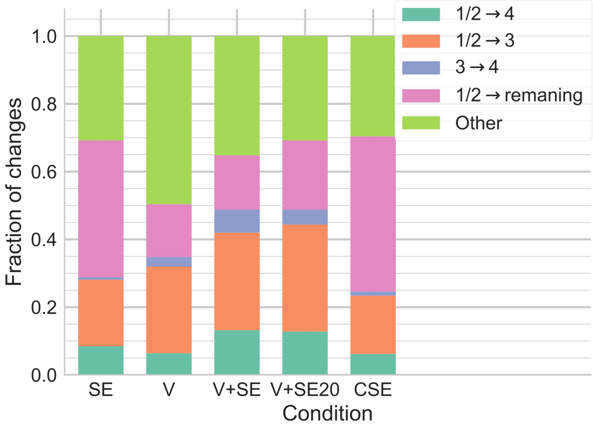


Figure 5.3: Overview of vocabulary knowledge state changes aggregated across all topics for each condition.

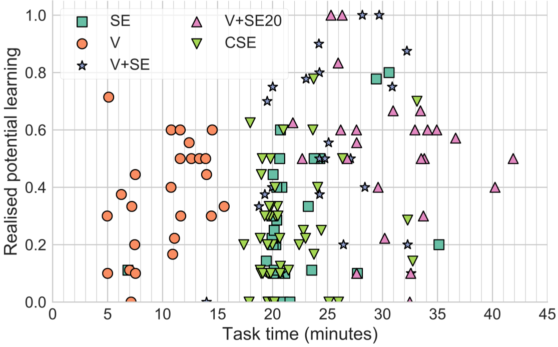


Figure 5.4: Active task time. Each point is a study participant.

video+search conditions **V+SE** and **V+SE20** this percentage is considerably lower (15.6%, 20.4% and 16% respectively), as all tested vocabulary items are mentioned in the video.

One expected difference between the **SE** and **V** conditions is the amount of time it takes to complete the task. As seen in Table 5.2 our selected videos have a length between five and fourteen minutes. As in any standard Web video player, our participants are free to pause, re-wind and skip ahead. In the **SE** condition, we require our participants to spend at least twenty minutes within our search system. We next examine the *actual* time spent on the respective interface(s). For **V**, this is the time difference between the first video-play event and the last video-stop event. For **SE** we consider this to be the time difference between the end of the interactive guide through the search interface and the time of the last submitted query, viewed document or mouse hover over a snippet (whatever came last). For the mixed video and search conditions we add up the times spent on those two interfaces. Figure 5.4 shows the relation between *RPL* and the active task time. Participants in the video-only condition spent between five and sixteen minutes (median 10m55s), while most of our **SE** participants spent the required minimum amount of time *actively* searching. Since we pre-set a minimum task time we have to leave an investigation into the *minimum* time required to achieve the same knowledge gain as participants in the video-only condition for future work. What we can say though is that on average **SE** participants require ten more minutes to achieve a comparable learning gain. This is clearly an *upper bound* as we did not investigate a reduction in search time.

Lastly, in Table 5.4 we list some key characteristics of the search session behaviours across conditions. Participants' search behaviour in **SE**, **V+SE** and **V+SE20** was very similar—participants are unlikely to noticeably change their search behaviour considerably within a single session. Prior work [45] has shown that within-session learning is possible, however, a large user population is required in order to observe the small changes in behaviour reliably. Most of our participants submitted between six and ten queries during the search session (four example sessions of our participants are shown in Table 5.5) that were 3-4 terms long, which is typical to current Web search queries. On average participants in the **SE** condition clicked on 10 links per session, mostly within the Web vertical. They also bookmarked slightly more documents than they clicked (on average 12.5) and spent on average nearly 7 minutes reading the clicked documents. These numbers indicate that our crowd-workers engaged with our search system and the task at hand as intended.

Table 5.3: Learning effectiveness metrics. Superscript ^{X} indicates a statistically significantly higher metric than condition X (Kruskal-Wallis, † p-value < 0.05, ‡ p-value < 0.01).

	SE	**V**	**V+SE**	**V+SE20**	**CSE**
ALG	0.281	0.320 CSE†	0.420 SE† CSE‡	0.444 SE‡ V‡ CSE‡	0.234
RPL	0.296	0.368 CSE‡	0.501 SE† CSE‡	0.518 SE‡ V† CSE‡	0.254

5.4.2 Instructor-Designed Learning with(out) Search Support

To address **RQ3.2**, we now explore whether a search phase immediately following the instructor-designed learning phase has a significant impact on the learning gain. The results in Table 5.3 indicate that this is indeed the case: both metrics *ALG* and *RPL* increase for **V+SE** and **V+SE20** compared to condition **V**. Both video and search conditions lead to significantly higher learning gains (absolute and potential) than the search-only condition; **V+SE20** significantly outperforms **V** as well. While in the video-only condition participants are able to increase their knowledge for slightly more than a third of vocabulary terms not known to them (RPL_V=0.37), in **V+SE** as well as **V+SE20** this is the case for more than half of the previously unknown (RPL_{V+SE}=0.5, RPL_{V+SE20}=0.52) vocabulary items.

Figure 5.3 shows two interesting insights on the knowledge state changes: first, **V+SE** and **V+SE20** participants are more certain about their learning than participants in the video-only condition (with the number of vocabulary state changes from 1/2⟶ 4 doubling); secondly, participants are able to confirm their partial knowledge to a higher degree—with most knowledge state change transitions of the type 3 ⟶ 4 occurring in the **V+SE** and **V+SE20** conditions. With respect to time-on-task (Table 5.4) our **V+SE20** participants spent on average just four more minutes searching than our **V+SE** participants, despite

Table 5.4: Basic search behaviour characteristics across the search conditions shown in Average (Standard Deviation). The clicks column lists the clicks on documents in the Web vertical as well as the aggregated clicks on all other verticals. For the Max. Clicked Rank column, we average the maximum click rank of each participant.

	Search Session Length	#Queries	Query Length	#Clicks	Max.	#Bookmarks	Reading Time
	[in minutes]		[in words]	Web vertical/ Other	Clicked Rank		[in minutes]
SE	21m58s (4m59s)	7.50 (6.12)	3.53 (2.14)	10.42 (6.96)/0.38 (1.16)	12.40 (11.67)	12.50 (8.03)	6m53s (5m08s)
V+SE	17m42s (11m13s)	8.16 (5.87)	3.12 (1.53)	12.44 (8.79)/0.56 (2.80)	10.04 (6.51)	10.12 (8.38)	7m23s (5m18s)
V+SE20	20m56s (2m36s)	7.48 (5.75)	3.31 (2.47)	13.20 (14.80)/0.68 (1.70)	10.78 (7.21)	14.72 (15.05)	6m37s (4m43s)
CSE	23m13s (12m52)	6.22 (3.86)	3.32 (2.06)	12.00 (6.89)/0.24 (0.55)	8.40 (8.57)	7.62 (7.88)	6m09s (3m32s)

Table 5.5: Four example search sessions logged in our experiment.

Water quality aspects	**Depression**	**Qubit**	**Anesthesia**
chemical processes relevant to ensure safe drinking water→threats to drinking water→chemical water treatment→safe drinking water supply→chemical water treatment	symptoms of depression→addressing depression→natural remedies for depression→natural remedies for depression websitetype:.org→natural remedies for depression .org→recognizing depresion symptoms→treatment for depression→causes of depression→supporting someone with depression	qubit→qubit transpose→quantum bits basics→quantum bits calculating length→quantum bits unitary vector→quantum bits amplitude→quantum bits amplitude vector→quantum bits basic state→quantum bits terminology→quantum bits calculating the transpose→quantum bits notation→quantum bits filetype:pdf	how anesthesia works → regional anesthesia → inhalational anesthesia → intravenous anesthesia → what does anesthesia do to your body

Table 5.6: Location of vocabulary items (VIs)

VKS change	Measure	SE	V+SE	V+SE20	CSE
1/2 →1/2	%clicked docs with VIs	9.68	1.68	2.01	18.12
	%snippets with VIs	0.12	0.01	0.06	0.19
1/2 → 3	%clicked docs with VIs	7.89	7.31	6.78	14.16
	%snippets with VIs	0.08	0.12	0.18	0.25
1/2 → 4	%clicked docs with VIs	7.26	4.52	2.46	6.59
	%snippets with VIs	0.11	0.09	0.05	0.15

the quite different minimum task times (for **V+SE** the total task time is set to 20 minutes, for **V+SE20** it is 20 minutes for just the search part); indeed Figure 5.4 shows that only a minority of participants quit the task immediately after reaching the minimum task time.

Lastly, we also consider to what extent the appearance of vocabulary items in the clicked documents and on the SERP (snippets) is indicative of knowledge state changes. This investigation is inspired by the observations reported in [45] where users were found to draw terms from the SERP and viewed documents to formulate subsequent queries within a search session. We here focus on the absence or presence of the tested vocabulary items within the viewed documents and the SERP and bin the vocabulary items according to the knowledge state change they underwent. The results are shown in Table 5.6. Although one might expect a particular trend (the more often a vocabulary item appears in the viewed documents, the higher the knowledge gain), there is actually none across the single-user search conditions; presence or absence of vocabulary items is not sufficient to approximate knowledge gains. Even more surprisingly, in the **CSE** condition we observe the opposite: for vocabulary items that remain largely unknown a larger percentage of documents contain those terms than for vocabulary items our participants increased their knowledge on. These two results point to the fact that a valid proxy of learning needs to measure much more than term absence/occurrence.

5.4.3 Collaborative Search As Learning

With **RQ3.3** we aim to explore whether our instance of a search as learning task can benefit from users collaborating together. Table 5.3 shows that in contrast to our hypothesis, CSE does not lead to increased learning gains compared to the other conditions. On the contrary, we observe our CSE participants to perform significantly worse than participants in the video-only condition (*ALG* of 0.32 vs. 0.23, i.e. **CSE** participants learn one word less than **V** participants on average) as well as both video+search conditions. One explanation can be found in the fact that despite spending more time within our search system than participants in all other search conditions, time is spent on the collaboration process—time that is not spent reading documents, as evident in Table 5.5, where **CSE** participants have the lowest average reading time compared to all other conditions. The median number of chat messages collaborative pairs wrote was 13 (minimum 1, maximum 54) and of all clicks, 6.15% came from entries in the query history and bookmarking widget that their partners made. As backed up by prior works on CSE systems, collaborators are efficient at sharing the bookmarking work: the average number of bookmarks per participant is indeed the lowest in the **CSE** condition, as participants here have a partner to contribute bookmarks as well.

Figure 5.5 provides us with another interesting insight with respect to our participants' academic background (high school certificate, undergraduate degree and graduate degree); here, the spread of realised learning potential within each background and condition is shown. Participants with an undergraduate degree show consistently higher gains than participants with a high school certificate (across all conditions the median realised learning potential is higher for undergraduates). Surprisingly, participants with self-reported graduate degrees do not follow this trend consistently, they perform especially poorly in the **SE** condition. The participants within the **CSE** condition show very similar learning potential (more so than participants in other conditions), likely due to the fact that the pairing of participants was random, instead of being based on a shared academic level. We also find that the spread in realised learning potential is small, there are few positive outliers.

Given the comparably low learning gains in the CSE condition, we explored whether our participants experienced particular difficulties finding information during the task, a question we included in the post-test.

Table 5.7 shows the result of an open card-sort approach (here, we merged the two search+video conditions); two of the authors independently sorted the 126 open answers submitted for this question in the into groups, discussed differences and then created a composite of the two results. We found seven categories of difficulties. Surprisingly, in the **CSE** condition nearly half of the participants (48%) indicated to not have encountered any difficulties, a higher level of satisfaction with our search system than participants in the single-user search conditions (where 38% reported no issues).

Interestingly, while in the **CSE** and **V+SE(20)** conditions, searching for the right information was the most often reported difficulty (e.g., "*No difficulty in finding information, however most websites gave more of an overview around physical treatment methods and did not use technical language or use the terms used in the lecture.*"), in the **SE** condition, where participants neither had a lecture video as a basis, nor a partner to exchange information with, this issue was only mentioned 12% of the time. Thus, the participants that achieved the highest learning gains overall, self-reported the largest difficulties with the search phase.

Table 5.7: Overview of participants' self-reported difficulties.

Category	**SE**	**V+SE/V+SE20**	**CSE**
No problems reported	38%	38%	48%
Task setup	19%	2%	—
Unclear focus	12%	8%	—
Searching	12%	40%	32%
Sensemaking	8%	10%	12%
Credibility of sources	7%	—	8%
Attitude	4%	—	—
Video >> search	—	2%	—

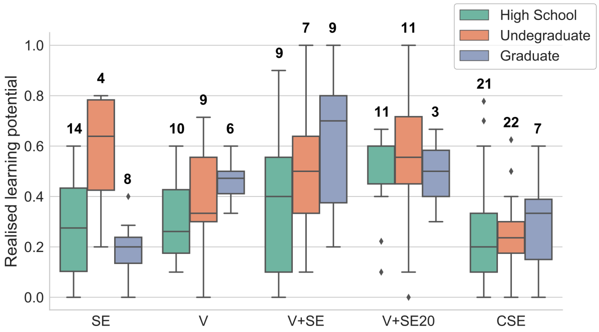


Figure 5.5: Relationship between highest academic degree and realised learning potential. The number of participants represented by each box is shown.

5.5 Limitations

The limitations of this chapter can be described in three folds: (i) our measurements of learning gains; (ii) our interface was not designed for learning; and (iii) we paid crowdworkers to conduct a learning task. First, we proposed a dataset that, even though it was intended to be very realistic, the type of learning task and the vocabulary we used were limited. In particular, we used graduates in the Computer Science department to rate the vocabulary difficulty of general topics. We acknowledge that this could have biased our vocabulary set, but this also leaves space for future work. Second, we acknowledge that we did not design our interface for learning. Consequently, our participants may have had difficulty using the task for learning. Lastly, we paid crowdworkers to conduct a learning task, which may not be realistic as a learning setting. We acknowledge that learning can be affected by other factors such as environment and interest. Paying crowdworkers could have a higher impact on learning compared to a real classroom scenario.

5.6 Conclusions

Is (individual and collaborative) search a viable alternative to instructor-designed learning? This is the question we set out to answer in this chapter. To this end, we carefully iterated over and designed a crowd-sourced user study with five different conditions that pitted instructor-designed learning (in the form of short high-quality video lecture as commonly found in online learning platforms) against three forms of search: (i) search-only, (ii) search as a support tool for instructor-designed learning and (iii) CSE. 151 participants spent a total of 138 hours in our experiment; they posed 897 queries and clicked on 1,512 documents. We measured their learning gains in a vocabulary learning task, which, though testing only lower cognitive skill levels [4], proved to be sufficiently challenging—not just for crowd-workers without a higher degree, but also for crowd-workers with a tertiary education (Figure 5.5).

We made a number of important findings: (1) participants in the instructor-designed learning condition reached up to 24% higher learning gains (measured in *RPL*) than participants in the search-only condition; (2) instructor-designed learning supported by search is superior to instructor-designed learning alone, leading to a 41% increase in realised potential learning; at the same time, these increases in learning gain do not translate into a higher confidence in the search process; (3) in our short-term learning task (approx. 20 minutes), CSE is not competitive as the collaborative overhead leaves less time for the retrieval and sensemaking steps of the search process: the learning gains decreased significantly compared to the video conditions. Considering that we pitted very high-quality lecture videos against search, we consider these results as an indication that search can be a viable (though worse) alternative to instructor designed learning, especially in situations where no high-quality video material is available.

Following the work in this chapter, several studies have tackled a number of limitations of our study:

- We have restricted ourselves to measuring learning only at the end of the search task. Roy et al. [119] extended the work in this chapter and presented a user studied to understand when learning occurs during a search session by measuring learning gains in regular intervals. Recall that in this chapter, measured learning gains via a vocabulary learning task. Roy et al. [120] and Roy et al. [121] considered the effects of different highlighting and note-taking strategies on learning-oriented search tasks and measured not only based on a vocabulary learning task but also essay writing task. A line of future work is how to capture learning gains via automated metrics. All the three works [119–121] used and extended SearchX presented in Chapter 2.

- To explore the benefits of interface elements that support learning, Câmara et al. [28] investigated an instructional scaffolding widget by extending SearchX, which guided participants in their learning during the search task. In instructional scaffolding instructor or teacher teaches learners through various means for them to achieve their learning goals. Here the system acted as the instructor and provided a widget to guide the learning of users during the session. They found that scaffolding methods did not significantly improve learning gain, but they impacted search behaviour considerably. An open question that still remains is whether more complex interface elements can better support tasks with increased cognitive skill levels.

- In our study, we designed the experiment to be completed by crowdworkers, which naturally restricted the possible task duration. Syed and Collins-Thompson [143] look at the long-term learning retention by conducting longitudinal study and evaluating learning gains with a repeated vocabulary task performed nine months later. Another line of future work is to continue exploring a more longitudinal setup. For instance, with MOOC learners would enable to conduct a large-scale study with a more significant number of conditions.

- Many other factors could impact the overhead of collaboration in the search for learning together. For instance, Xu et al. [163] investigated how demographics and traits of users could influence the user knowledge gains in a collaborative setting. In particular, they focused on patterns that aided collaborative learning, such as communication and division of labour. However, a similar limitation of their work and our work is that they only conducted the collaborative setting in pairs. Therefore, there is room for research in the direction of CSE for learning.

6

Exploring Collaborative Navigation Support in Collaborative Product Search

In this chapter, we investigate the benefits of collaborative shopping-oriented search systems. In particular, we aim to examine the benefits of enabling and increasing sense of co-presence in collaborative online shopping and how it can improve group coordination during the decision making process of purchasing a common product. In particular, we aim to analyse the five stages of group decision making process: (i) recognise a collaborative shopping need in a group; (ii) search and collect products from the online catalogue using a CSE system; (iii) evaluate the collected products; (iv) make a purchase decision; and (v) post-evaluate a purchase decision. One concrete manner to increase the sense of co-presence is collaborative navigation. Collaborative navigation is the process of allowing a group to navigate and share information via an intermediate interface. In our work, we investigate to what extent a group of collaborating users differ in their search and decision making behaviour when faced with either separate or shared navigation features (the latter being a specific type of collaborative navigation). To this end, we conducted a user study with 30 groups given a collaborative shopping task using either separate or shared navigation. We find that shared navigation participants were more effective in their search space exploration, and navigation support did not significantly impact the purchase and post-purchase decision making stages. Importantly, this is the first work to shed light on the impact of navigation support in collaborative search on the five stages of a group decision making process, with substantial design implications for collaborative shopping-oriented search systems. The contributions of this chapter have been published in [98].

6.1 Introduction

Recently we have witnessed the emergence of online collaborative tools (e.g., Google Docs, Slack, Zoom, etc.) that can assist a group of users in sharing and constructing knowledge, communicating with each other, and coordinating their work. In the e-commerce domain, online retailers have begun exploring *collaborative online shopping* tools to provide a group of customers an experience closer to physical shopping in the real-world—and consequently improve engagement with the retail website. For instance, *Amazon*[1] offers its customers the ability to create collaborative wish lists and chat on its website; in addition, households can conduct voice search and shop together using the `Alexa` voice assistant. In addition, the COVID-19 pandemic has prompted a rapid change in how customers shop. *Squadded*[2] has launched the *"Shopping Party"* browser extension that allows users to shop virtually together with friends on fashion retailer websites. By enabling social engagement between collaborators on their website, retailers may observe business advantages such as increasing time on site, product views, and intention to return [67, 74].

Specifically, in a collaborative online shopping experience, a group of users typically go through five stages of group decision-making [60, 75]. First, the group needs to recognise a collaborative shopping need, which can include one or more people advising another member on a product to buy or the whole group buying a product together. Stage two consists of searching and collecting products that are then evaluated in the third stage. The first three stages can be complex as customers need to process a variety of sources of information such as customer-generated information and seller information and e-commerce websites need to retrieve from these sources [150]. In the fourth stage, a purchase decision is made, and lastly, in stage five, the group evaluates their perceived satisfaction with their purchase decision. Besides these stages, phases of discussion occur in all of the stages and serve as a moment of regaining group coordination. A number of researchers have investigated how different collaborative features can support or hinder groups when working through a collaborative shopping task [33, 49, 156, 166, 170].

A number of collaborative shopping systems have been developed [49, 105, 166, 170] in order to support a group of users to search together for products, explore product recommendations, and make a purchase decision. These systems have been studied with various approaches such as product display format, recommendation agent, and chatting tool [65], to enhance remote[3] customers' sense of co-presence, improve shared understanding, reduce conflict resolution, and increase intention to return to the retail website. In particular, enabling and increasing a *sense of co-presence* has been shown to be an essential aspect in collaborative shopping as it can improve group coordination during the decision-making process [166]. Increasing awareness has been found to improve the sense of co-presence of collaborators and consequently to benefit the collaboration [127]. One concrete manner to increase the sense of co-presence is through *collaborative navigation.*

Collaborative navigation is the process of allowing groups to navigate and share information via an intermediate interface [125]. Different types of collaborative navigation exist, such as *split screen* [166] and *shared navigation* [170]. In the literature, there are numerous examples of systems that provide such functionality [33, 166, 170]. Collaborative

[1] `https://www.amazon.com/hz/wishlist/intro` — all URLs last accessed May 24th, 2021.

[2] `https://www.squadded.co/`

[3] Customers that are synchronously remotely located on different devices

shopping systems often support users when navigating an e-commerce site by sharing a browser's session, which can make members in a group aware of the location of each other—or observe another member's shopping session [166]. Collaborative search (CSE) systems have also been developed with navigation features in mind to improve the awareness of members, their coordination, and the sharing of knowledge [101].

Although CSE systems are not often connected to collaborative shopping systems in the literature, both kinds of systems provide similar functionality to aid in a collaborative scenario.[4] Specifically, navigation support features determine how a user employs a CSE system together with other users. For instance, in a *separate navigation features*[5] setup (which we refer to as `SepNav`), users conduct their own searches and can act independently of other users. In contrast, in a *shared navigation features* setup (which we refer to as `SharedNav`), any action taken by a user such as submitting a query or clicking on a result may change the interface other members of the collaborating group currently see. While not common, shared navigation features have been shown to reduce conflicts between group members and make users more effective in their discussion [170]. However, previous works did not focus on understanding the differences in group behaviour of navigation support features that can influence the entire group decision making process [33, 156, 166, 170].

In this chapter, we investigate how the different stages of group decision making are impacted by the type of navigation support (`SepNav` vs. `SharedNav`) in a collaborative product search task. To this end, we designed a collaborative shopping system that consists of a CSE system, with a product catalogue as an underlying corpus. User interface features that users have come to expect from online shopping sites (such as faceted search) were included. Through a user study with $N = 64$ participants, we perform a comprehensive analysis of the effect of the navigational mechanisms during the group decision making process to answer the following research question (**RQ4**): *How and when do navigational mechanisms impact the group decision making process in collaborative online shopping tasks?* We find that `SharedNav` participants were more effective in their search space exploration, and we find no significant impact of navigation mechanisms on the purchasing decisions and post-purchase decision evaluation stages of participants.

The main findings of this chapter are:

- We find that groups in the shared navigation were more effective in their search space exploration.
- We find that shared navigation causes less usage of typical CSE features.
- Importantly, we do not find a significant impact of shared navigation on the purchase decision and post-purchase decision perceived satisfaction.

6.2 Related Work

In Chapter 2 we have provided details in CSE in general, in this section, we look closely at collaborative navigation, collaborative shopping, and group decision making.

[4] Hereon, we refer to both types of systems as *CSE systems*.

[5] Features that do not interfere with user navigation, such as the mouse location of other participants, or widget displaying other participants recent queries.

6.2.1 Collaborative Navigation

Collaborative navigation is a feature that determines how users navigate the functionalities of a collaborative system. Collaborative browsers are one type of system that can typically support synchronous collaboration via yoked views, where a user's interactions causes the browsers of other group members to execute the same interaction automatically. For instance, recent technologies such as TogetherJS[6] provide web application developers with the ability to let users initiate a shared session on their website—alongside chat capabilities (through text or video/audio). Screen sharing solutions also support collaborative navigation. However, users involved can only share their screens—others have no ability to collaboratively control what is happening [147].

In the context of *CSE*, collaborative browsers and screen sharing solutions can support a group collaboratively searching. However, these browser solutions do not contain search-specific features. In contrast, CoSearch [3] was implemented to provide multi-device collaborative navigation to support *co-located*[7] CSE and giving each user a mobile device, centralised by a desktop computer. Similarly, WebSurface [151] and We-Search [104] provide co-located users a common display that users can navigate in their CSE sessions.

The literature also provides examples of remote collaborative navigation to support CSE. SearchTogether is an example of a system where users can split their search results by opening them in a new browser tab, and sharing this with others. This is known as *separate* or *independent navigation*, where users in a collaborative group use available collaborative features independent of one another. This approach has been widely used by researchers in the area [109, 115, 128]. In contrast, *shared navigation* mimics collaborative browsing features, where users can see (in real-time) the interactions that other users make, such as mouse clicks, scrolls, or issued queries [33, 156, 166, 170]. However, *shared navigation* features may introduce competition in attention and resources, potentially distracting members from the interactions of others, and may *'fight'* for the control of the task or conversation [62]. This can have a negative effect on group coordination and require additional effort in conflict resolution [170]. To overcome this potential side effect of shared navigation, [166] introduced another form of collaborative navigation support called *split screen navigation*. Here, dyads (pairs) can see the navigation of their partner in real time. Through a user study, the authors found that *shared navigation* best improves navigation in terms of alignments of what collaborators communicated, and what they were seeing (via an eye-tracker), but *split screen navigation* encourages more diverse searches (i.e., a better exploration of the search space). Our work builds on [166, 170]. Specifically, we extended SearchX to support *shared navigation* and to support a collaborative shopping experience. In contrast to [166, 170], we restrict *shared navigation* to the SERP only. We argue that this limited approach reduces the likelihood of user confusion, and will reduce the cognitive load of using the interface.

[6] https://togetherjs.com/

[7] One CSE system design dimension is *location* [52]: remote collaboration means that users are not co-located, and thus their coordination requires a means of communication such as text chat or video/voice chat. In contrast, co-located collaboration means that users are physically in the same space.

6.2.2 Collaborative Shopping

Morris [100] observed that 25.7% of the participants in their 2008 survey on collaborative web search had collaborative shopping needs. In the meantime, tech start-ups have incorporated social presence in their online platform [8]. Although online shopping may not always involve search episodes, searching is an important part of the shopping journey, and facilitating a better search experience can impact customer experience [150]. A few works in collaborative shopping investigated the design and evaluation of collaborative features that can improve a collaborator's experience of co-presence and communication [74, 133, 156, 166, 170]. Co-presence support tries to increase collaborators awareness of each other in order to improve the social presence while shopping together. This has been achieved by developing system design components with collaborative navigation support [156, 166, 170] and embodiment [74]. Communication support or media richness has been studied as a component to support co-presence, but also to support coordination [133, 170].

Although many of the studies described above discussed how co-presence support can enhance collaborative shopping experiences, none have investigated how searching impacts the collaborative shopping experience. To bridge this gap, [49] developed `ShopWithMe!`, a collaborative shopping and searching system, and conducted a user study to investigate whether CSE tools can support collaborative shopping. They found that CSE improves information sharing among group members collaboratively shopping for experience goods. Our own work follows this line of inquiry. However, we aim to investigate *how collaborative navigation support can impact users in the group decision making process*, which also includes searching together as one of the stages in the process.

6.2.3 Group Decision Making

In the e-commerce environment, decision making has been extensively studied—focusing on how decision aids can improve the decision making process of customers [60]. Kim and Srivastava [75] described in general the customer's decision making process in e-commerce, and examined how social influence affects each decision making process stages. They argued that during a shopping episode, a customer typically goes over *five stages of the purchasing decision making process*. **Stage 1**: recognise their information need (e.g., defining product requirements such as price, colour, size). **Stage 2**: conduct exploratory searches of the product catalogue and collect the candidate products. **Stage 3**. evaluate and compare the candidate products. **Stage 4**: make a purchase decision. **Stage 5**: evaluate their post-purchase decision.

In addition, communication has been found to play a key role in group decision making. Nakamura et al. [105] also investigated group decision making in CSE in a restaurant booking task. Based on the conducted user study with 24 participants, they found that communication happens during the entire decision making task—with a leaning to more communication at the beginning and the end of the task. Given that very few works in CSE research studied the impact of CSE tools on decision making [61], we want to understand what impact CSE features such as navigation support have on group decision making in collaborative shopping.

[8] https://www.ecommercetimes.com/story/78142.html

6.3 Experimental Design

Now we present our experimental design of this chapter. We introduce the SearchX setup, the experimental conditions, tasks, experimental procedures, and evaluation measures.

6.3.1 SearchX Setup

We used our search system SearchX version V4 (see Figure 2.3 in Chapter 2). Concretely, we provided our participants with the SearchX's UI depicted in Figure 2.9 with a search box ① with search facets ⑰, search filters ⑬, a document (product) viewer (see Figure 2.8) without document rating ⑨ and document annotation ⑩, recent queries ④, saved items ⑲, shopping basket ⑳, and group chat ⑥.

Data Collection We utilised the *Amazon* public dataset with six different product domains (Amazon departments) (see Chapter 2.3.4 for more details). Concretely, we indexed the collection using Elastisearch and used multi-field BM25 to provide search results and facets to the user.

Shared vs. Separate Navigation

The SepNav condition is the baseline condition: participants of a collaborating group each issue their own queries, and view only their own search results on the SERP. This means that participants subjected to this condition did not see the popup ⑭ in Figure 2.3.

In contrast, in the SharedNav condition, participants within the group had their SERPs synchronised once one of three actions occurred: (i) any member of the group issued a query; (ii) filtered the search results; or (iii) changed the results page (via the pagination component). Our approach was inspired by [170] and [166]. Our study, however, does differ in relation to SERP synchronisation, and would only occur if one of the three actions listed above occurred. Other interactions with the user interface, such as scrolling, were not synchronised. Thus, our participants still had the freedom to explore a limited set of search results independently. We chose this setup to limit user confusion as in literature absolute synchronisation turned to more uncoupling events between two collaborators [170]. Besides, as we conduct our experiment with more than two collaborators, requiring our participants to use two or more displays would be not feasible. In other words, our implementation can be scaled up to more than two collaborators without any additional requirements.

The SepNav condition has been trialled not only as a baseline in collaborative shopping research [33, 49, 105, 170], but also in numerous studies in collaborative travelling planing, collaborative-work tasks, and collaborative search as learning [29, 95, 96, 102, 105, 109].

6.3.2 Experimental Conditions

As outlined in Section 6.2, a number of experimental conditions have been trialled in previous works regarding navigational support in collaborative (shopping) systems. Importantly, our work aims not to compare novel or reproduce navigation support approaches, *but to understand how and when navigation support approaches affect the group decision-making process*. We randomly assigned participants to one of two conditions.

`SepNav` In this condition whenever one participant in a group searches, the SERP is updated only for the participant that posed a query.

`SharedNav` In this condition, whenever one participant in a group searches, the SERP is synchronised among all participants of the group.

Participants under both conditions used the interface as outlined in Chapter 2.3; those assigned to `SepNav` however would not see the callout 14 in Figure 2.3; participants worked independently.

6.3.3 Training and Main Task

We designed our study tasks intending to make them complex and nuanced enough to require collaboration to accomplish the shopping task [27]. In particular, based on observations in the literature, a shopping task could result in individualised searching behaviour rather than collaboration and collaborative sense making [140]. Hence, with collaboration as a focal point, participants completed one training task, and one main task.

Training Task In this task, participants had 15 minutes to search for an everyday product that their employer will buy for each of them (the *same* product) to enhance working from home productivity. This task aims to familiarise participants with the CSE system, and to search together. The participants went over the first four stages of the group decision task outlined by [75]. We did not require them to fill in a post-purchase decision questionnaire.

Main Study Task Figure 6.1 presents the template we used for the main task. This is a single task and each group conducted a single task as described Figure 6.1. The goal of this task is to elicit a collaborative shopping task so that we can observe the five stages of the group decision making process.

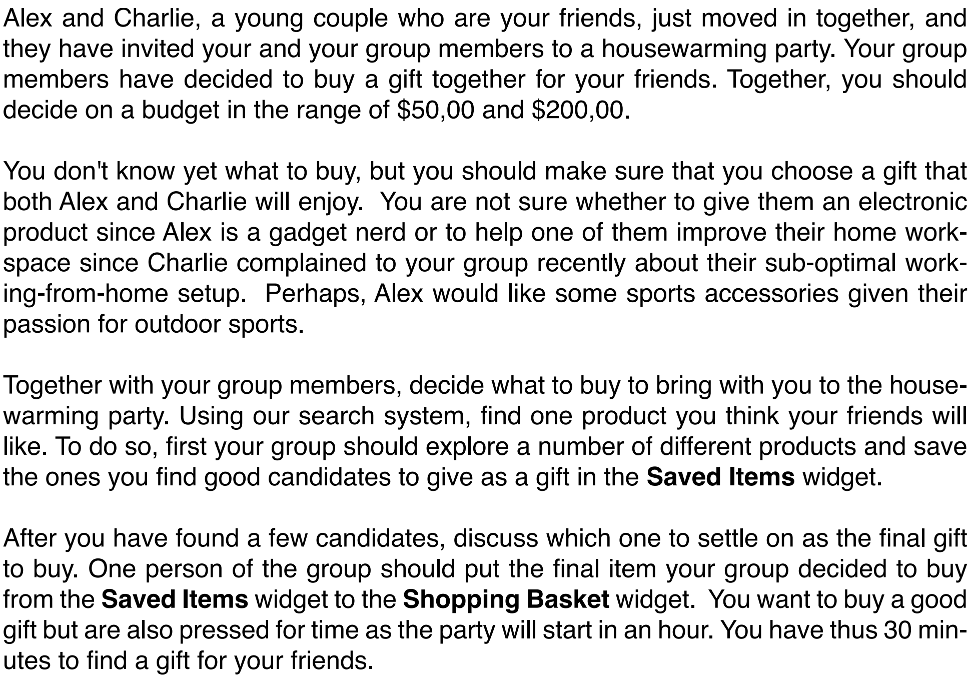
6.3.4 Experimental Procedure

We scheduled an appointment for an online lab experiment[9] with each group of participants at a time they were online together—preferably in different locations.[10] Once participants had accepted a consent form, they were redirected to our server with the instance of `SearchX` running for their assigned condition. An overview of the experimental procedure is shown in Figure 6.2. We now outline each of the steps.

1. The participants met up in a video call with the experimenter dispensing initial instructions. Participants were provided with a link to the experiment, after which the video call ended.
2. The experiment began with participants independently answering pre-task questions.
3. Next, participants were moved to an online *waiting room*, where they waited for other participants that signed up with them to finish the pre-task questions.
4. Once the participants moved to the CSE phase, they first undertook a short, automated, interactive guide of the `SearchX` interface. The interface's key features were highlighted to the participants, along with a short explanation of what the highlighted components

[9] Due to COVID-19 regulations, we could not conduct a physical lab experiment.

[10] In case participants lived together and were in the same room, we asked them to not talk to each other during the experiment.



Alex and Charlie, a young couple who are your friends, just moved in together, and they have invited your and your group members to a housewarming party. Your group members have decided to buy a gift together for your friends. Together, you should decide on a budget in the range of $50,00 and $200,00.

You don't know yet what to buy, but you should make sure that you choose a gift that both Alex and Charlie will enjoy. You are not sure whether to give them an electronic product since Alex is a gadget nerd or to help one of them improve their home workspace since Charlie complained to your group recently about their sub-optimal working-from-home setup. Perhaps, Alex would like some sports accessories given their passion for outdoor sports.

Together with your group members, decide what to buy to bring with you to the housewarming party. Using our search system, find one product you think your friends will like. To do so, first your group should explore a number of different products and save the ones you find good candidates to give as a gift in the **Saved Items** widget.

After you have found a few candidates, discuss which one to settle on as the final gift to buy. One person of the group should put the final item your group decided to buy from the **Saved Items** widget to the **Shopping Basket** widget. You want to buy a good gift but are also pressed for time as the party will start in an hour. You have thus 30 minutes to find a gift for your friends.

Figure 6.1: The task information template, as presented to participants of the study. This is for the main study task only.

do. Participants were then given the training task in order to familiarise themselves with the collaborative system. After 15 minutes, we automatically moved the participants to the main task.

5. Participants read the instructions for the main task, before undertaking it with their team member(s). Our participants had 30 minutes to complete the task that is outlined in Figure 6.1.
6. Once the main task was completed, we then provided a final, post-task questionnaire. This was completed individually.

6.3.5 Pre-Task and Post-Task Questionnaires

Besides six demographic questions, our pre-task questionnaire included six CSE questions as presented by [101] to aid priming of the participants for the upcoming CSE tasks. We included a short explanation of CSE and when and how it can happen, along with three images of groups collaborating (co-located and remote collaboration). These questions were also included in Chapters 3, 4, and 5.

We provided a post-task questionnaire with 15 questions to capture each participant's *experiences* of shopping collaboratively with the provided system listed in Table 6.1. Participants were able to indicate their perceived satisfaction with the group purchase decision, their opinions on the search and discussion phases that took place, their participation in the group decision, and their perceived difficulty of the group decision-making process.

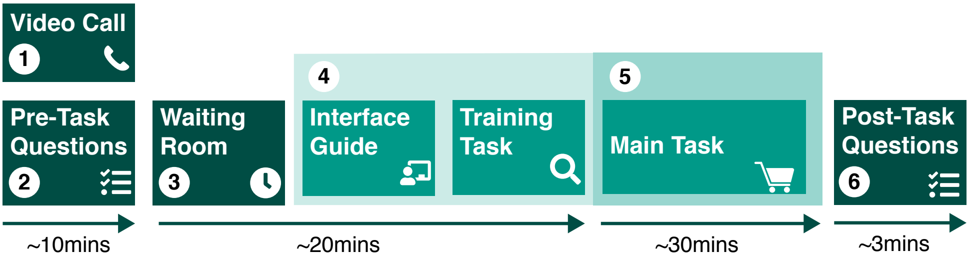


Figure 6.2: The experiment workflow for participants of this study. For more detail on the six main steps (as numbered on the diagram), refer to §6.3.4.

Answers were provided using a seven point Likert-scale; an open-ended question was also asked soliciting for their thoughts on the experience.

6.3.6 Study Participants

Over the course of 21 days, a total of 64 study participants completed our experiment successfully across the two experimental conditions. We recruited our participants from the *Delft University of Technology*, the Netherlands, via social media and internal communication channels. In total, 30 groups signed up in groups of two (26 groups) and three (four groups). We randomly allocated 13 groups of size two and 2 groups of size three to each experimental condition. Group members knew each other before the experiment. Their median age was 23 (min = 14 and max = 49). 32 of our participants identified as female, 30 as male, and two as non-binary. Our participants come from various nationalities: mainly India (17) and the Netherlands (16); the remaining participants are from EU countries (15), Mexico (3), Brazil (2), USA (2), Turkey (2), Nepal (2), Pakistan (2), China (2), and Indonesia (1). Most of our participants reported to be advanced English speakers. Their academic backgrounds varied: 31 reported possessing an undergraduate degree diploma as their highest academic degree; 22 a graduate degree; and and the remaining a high school diploma. We paid each participant a gift card valued at €12, which is comparable with the Netherlands hourly minimum wage. The median time they spent in our experiment was 60 minutes including pre-questionnaire (10 minutes) and post-questionnaire 3 minutes and 7 seconds.

6.3.7 Evaluation Measures

To evaluate the impact of collaborative navigation support in the group decision making process, we captured a variety of events triggered by participants as they performed the product search tasks. Table 6.2 shows the events aggregated by event type. In our data analysis we aggregated the events in four types to capture the first four stages of the group decision making process. More precisely, Search events were used to serve as a proxy for **Stage 1** (information need recognition) and **Stage 2** (information search). Review events allow us analyse **Stage 3** (evaluation of selected products) and Purchase events **Stage 4** (purchase decision). We also analyse how communication takes place during the decision process with Chat events.

Pre-Task Questions	
Demographics	What is your highest academic degree so far? What is your age? What is your gender? Are you an English native speaker?
Online shopping experience	How often do you buy products online? How often do you buy products online together with someone else (friend, family member, partner)? What types of products/services do you buy online? Select the product categories that you often buy online Which information about the product you find helpful during your online shopping?
Post-task Questions	
System experience	It was easy to understand why products were retrieved in response to my queries. I didn't notice any inconsistencies when I used the system.
Task experience	It was easy to determine if a product was relevant to the task. How difficult was this task?
Collaboration experience	I could see what they others were searching and examining during the task. It was easy to share what I found useful during my searches with my partners. Did you find the collaborative features useful?
Post-purchase decision perceived satisfaction	Were you satisfied with the decisions that you group have made? Did you search products well enough to make a decision? Did you discuss with others well enough to make a decision? Did you expresss your opinions well enough in the discussion? Do you think the decisions were reflected by your opinion? What was your contribution during the task?

Table 6.1: Pre-task, post-task and after-task questions.

Table 6.2: Event types that were triggered by participants. Numbers refer to the system interface (see Figure 2.3).

Type	Events
Search	Issuing a query (via query box ① or recent query ④), paginate SERP ③, change facet ⑰, filter SERP ⑬, click on search results ⑭, save or add to item to the shopping basket ⑳
Review	Click on or delete saved item ⑲
Chat	Send a chat message ⑥
Purchase	Click on or delete shopping basket item ⑳

Based on the interactions we captured with events described in Table 6.2, we employed a range of evaluation measures commonly employed in *Interactive Information Retrieval* and collaborative shopping research [51, 96, 102, 166, 170].

We report aggregate search and review behaviours over a group 👥 and individually 👤. We report a number of behaviours associated with querying and interactions with the SERP such as product clicks, and time viewing products, for example. We also analysed the chat messages exchanged between participants; we report, among others, the time between messages and the balance of messages amongst group members. The latter was measured via the *Gini coefficient* [42] with the number of messages of each group member as input.

6.4 Results

We now address the results in light of our overarching **RQ4**. Significance testing between conditions was performed with the *Mann-Whitney U pairwise test* (at $\alpha = 0.05$), as we observed high variances for our measures. Note that ± values in the tables and corresponding narrative both indicate the *standard deviation*.

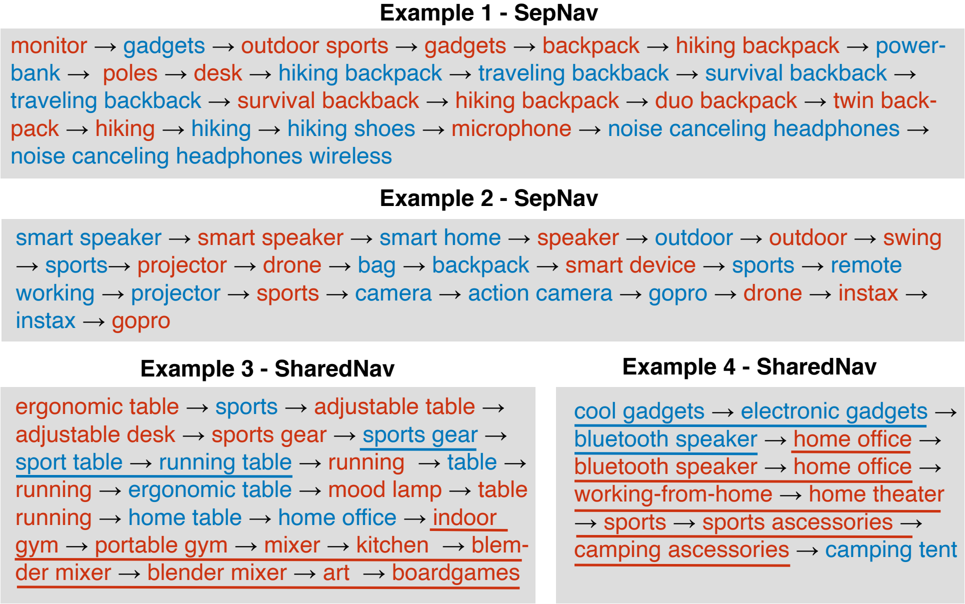
6.4.1 Descriptive Behaviour Analyses

Our **RQ4** considers *how and when navigational mechanisms impact the group decision making process in collaborative product search tasks*. We first focus on the *how* part of our **RQ4**. Table 6.3 presents an overview of our behavioural measures grouped by the four behaviour types: search, review, chat, and purchase. Results are reported across the two experimental conditions. For the first two columns of Table 6.3, mean values are the *average over the group* 👥 across each condition, and the last two columns *the average over participants* 👤 *across each condition*. We first examine the effects of collaborative navigation on the **Stage 1** (information need recognition) and **Stage 2** (product search and selection).

Our analyses show that for **Stage 1** (captured with behaviour measures **SE1–4**), we observe a significant difference in the number of queries issued via the search box (**SE1**) (10.44 ± 4.77 queries and 8.12 ± 4.79 queries for `SepNav` and `SharedNav`, respectively), and also via the recently issued queries component (**SE4**) (1.03 ± 1.89 queries and 0.22 ± 0.55 queries for `SepNav` and `SharedNav`, respectively) across *individual participants in each condition*. Additionally, we observe a significant difference in terms of the number of queries issued amongst group members measured as the Gini coefficient of the number of queries

Table 6.3: Mean (± standard deviations) of search, review, chat, purchase behavior metrics across all groups (👥) and participants (👤) in each condition. A dagger (†) next to SharedNav values denotes significance from the baseline, SepNav.

Stage and Measure			Over Groups		Over Participants	
			👥 SepNav	👥 SharedNav	👤 SepNav	👤 SharedNav
Search	SE1	Number of queries issued	22.27(±10.96)	17.33(±7.59)	10.44(±4.77)	8.12(±4.79)†
	SE2	Difference in number of queries among group members	0.08(±0.06)	0.15(±0.10)†	–	–
	SE3	Average query length (in words)	1.88(±0.57)	1.77(±0.39)	1.88(±0.61)	1.74(±0.44)
	SE4	Number of recent queries clicks	2.20(±3.10)	0.47(±0.74)	1.03(±1.89)	0.22(±0.55)†
	SE5	Maximum search pagination depth	4.35(±2.28)	3.45(±0.92)	4.23(±4.56)	3.18(±1.03)
	SE6	Number of filter selections	15.93(±9.04)	13.40(±8.40)	7.47(±5.87)	6.28(±5.33)
	SE7	Number of product clicks	9.53(±6.47)	9.07(±6.53)	4.47(±4.95)	4.25(±3.77)
	SE8	Number of saved products	10.73(±4.76)	9.07(±5.20)	5.03(±2.44)	4.25(±2.74)
	SE9	Number of unique product retrieved	698.13(±362.95)	465.73(±160.79)†	346.94(±194.15)	438.56(±149.05)†
	SE10	Average product view time (seconds)	11.73(±3.82)	10.98(±5.23)	10.54(±7.93)	9.83(±10.36)
Rev.	R1	Number of saved items click	18.40(±15.00)	9.40(±4.53)†	8.62(±6.89)	4.41(±3.30)†
	R2	Average saved product view time (seconds)	8.90(±5.49)	8.32(±5.32)	10.54(±7.93)	9.83(±10.36)
	R3	Number of deleted saved products	3.80(±5.45)	3.20(±5.17)	1.78(±3.23)	1.50(±3.03)
Chat	C1	Number of messages	70.20(±35.91)	74.40(±29.27)	32.91(±16.06)	34.88(±13.60)
	C2	Time between messages	25.69(±13.62)	21.96(±9.27)	56.78(±32.14)	48.67(±21.99)
	C3	Number of chat words	392.00(±214.54)	466.67(±210.20)	183.75(±104.03)	218.75(±114.68)
	C4	Number of chat clicks	23.47(±16.11)	19.47(±19.27)	11.00(±13.05)	9.12(±11.52)
	C5	Balance of messages among group	0.08(±0.05)	0.08(±0.05)	–	–
Purch.	P1	Number of basket products	1.93(±2.63)	1.87(±1.25)	0.91(±1.51)	0.88(±0.91)
	P2	Number of basket clicks	1.73(±3.37)	0.53(±0.74)	0.81(±1.97)	0.25(±0.51)
	P3	Number of deleted basket products	0.27(±0.59)	0.53(±1.13)	0.12(±0.34)	0.25(±0.57)
	P4	Time taken to add the last product to basket (mm:ss)	22:03 (± 5:15)	23:18 (± 4:49)	–	–



Example 1 - SepNav

monitor → gadgets → outdoor sports → gadgets → backpack → hiking backpack → powerbank → poles → desk → hiking backpack → traveling backback → survival backback → traveling backback → survival backback → hiking backpack → duo backpack → twin backpack → hiking → hiking → hiking shoes → microphone → noise canceling headphones → noise canceling headphones wireless

Example 2 - SepNav

smart speaker → smart speaker → smart home → speaker → outdoor → outdoor → swing → sports→ projector → drone → bag → backpack → smart device → sports → remote working → projector → sports → camera → action camera → gopro → drone → instax → instax → gopro

Example 3 - SharedNav

ergonomic table → sports → adjustable table → adjustable desk → sports gear → sports gear → sport table → running table → running → table → running → ergonomic table → mood lamp → table running → home table → home office → indoor gym → portable gym → mixer → kitchen → blemder mixer → blender mixer → art → boardgames

Example 4 - SharedNav

cool gadgets → electronic gadgets → bluetooth speaker → home office → bluetooth speaker → home office → working-from-home → home theater → sports → sports ascessories → camping ascessories → camping tent

Figure 6.3: Query logs over four individual search sessions, as captured in our experiment (with group sizes of two). Each colour represents a different participant. Underlined phrases are possible alternation of roles during the task.

of each participant (**SE2** in Table 6.3). This finding shows that often a group member in the SharedNav condition issued more queries than the others. This suggests that collaborative navigation instantiated with SharedNav features affected the capability of participants to formulate a larger set of queries, with one member taking the role of issuing the queries for the other.

Turning our attention to **Stage 2** (captured with behavioural measures **SE5–10**), we observe that the number of unique products retrieved (row **SE9** in Table 6.3) shows significant differences between the two conditions—both when examining results on a per group and individual basis. Groups in SharedNav navigation collaboratively covered a smaller range of the search space as participants are always brought back together. However, groups in SepNav *individually* covered a smaller range of the search space, as each participant retrieved a different number of products. This finding suggests that group members subjected to SepNav retrieved more previously observed results (from queries earlier in the session) than group members in SharedNav. To highlight this further, we showcase in Figure 6.3 the query logs from four of our sessions, two each from SepNav and SharedNav. We can see in Example 1 that there was an overlap of the search space by members of a group subjected to SepNav (i.e., hiking backpack is first explored by one member, and later the other member issues the same query, while the former continued in a different direction). Our finding is different to [166]—the authors *could not* find a significant difference in the number of retrieved items over a hotel search task between the SepNav and SharedNav conditions. This difference can be explained by the fact that for our study, only the SERP was synchronised for group members subjected to SharedNav—and not over actions such as scrolling, or viewing specifics about a product.

Next, we analyse the impact of collaborative navigation behaviours pertaining to how groups evaluated their selected products during **Stage 3**. This analysis is based on metrics **R1–3** in Table 6.3. We observe that group members in the SharedNav condition interacted with the saved items interface component (19 in Figure 2.9) significantly more often SepNav than SharedNav (18.40 ± 15.00 vs. 9.40 ± 4.53). A possible explanation for this phenomenon is that group members subjected to SepNav relied more on the saved items component to know which item the other member was searching for (and selecting) at the same point in time.

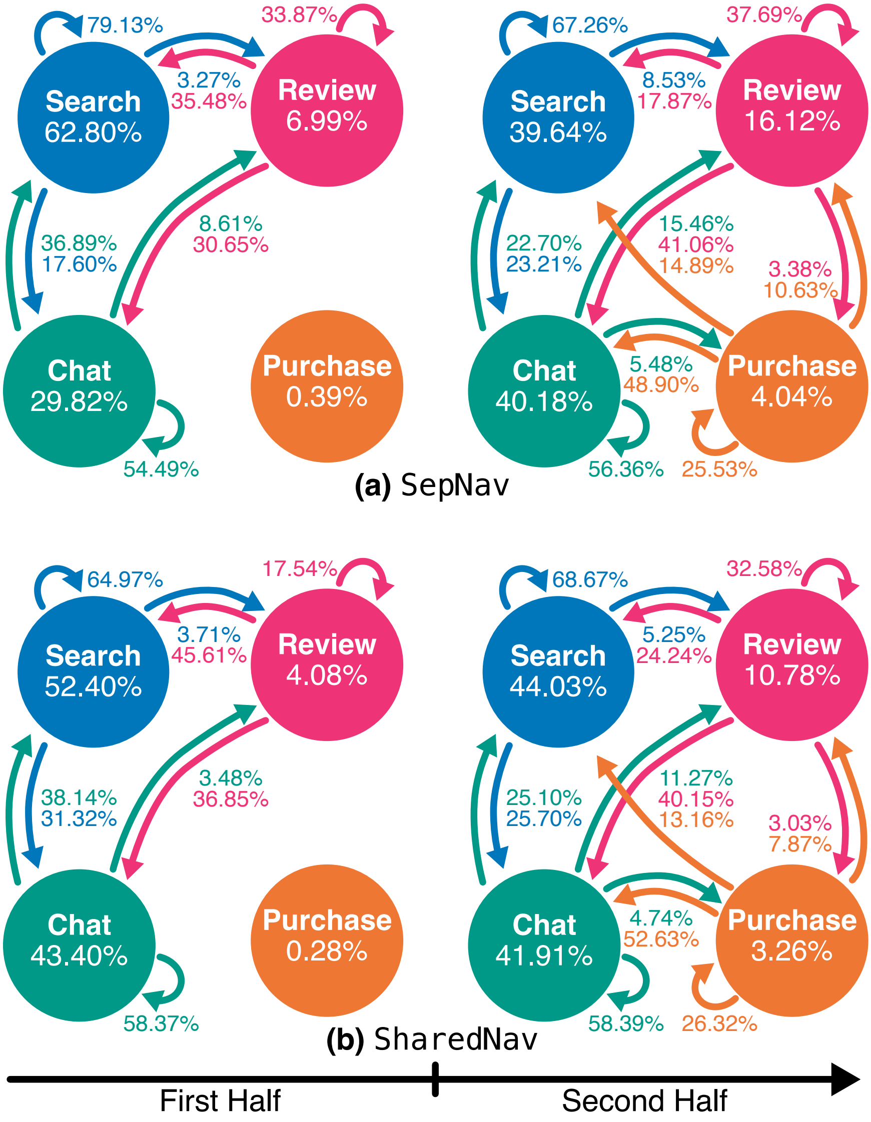
We now look at the behaviour metrics for **Stage 4**. Purchase behaviour metrics **P1–4** in Table 6.3 show that groups in both conditions purchased and analysed purchased products put in the shopping basket in a similar manner. From metric **P4**, we observe that groups finished adding products to the basket on average at times of 22:03±5:15 minutes and 23:18±4:49 minutes for the SepNav and SharedNav conditions respectively. As participants were given 30 minutes to perform the search task, agreeing on the shopping basket after 22 – 23 minutes on average indicates sufficient time was available for groups to make the decision without time pressure.

Lastly, from Table 6.3 (row **C1–5**), we analyse the phase of discussions that occurred within the chat messages. We observe no significant impact of collaborative navigation on the discussion in terms of summary metrics. Interestingly, despite having an unbalance in the number of queries among group members in the SharedNav condition, we did not observe an unbalancing of the number of messages exchanged (row **C5** in Table 6.3). One would expect that participants in SharedNav that issued queries less frequently during the task would communicate more to the participants issuing more queries. This suggests that there was possibly an alternation between roles during the collaborative shopping task—this can also be seen in the underlined phrases in examples 3 and 4 of Figure 6.3.

6.4.2 Time-based Behaviour Analysis

We now focus on the *when* part of our **RQ4**. In Section 6.3.7, we described the group decision making stages in sequence. However, group members could go back and forth in these stages. We zoom in on how this dynamics occurred along with the task. To do so, we show in Figure 6.4 a series of Markov models, where the stages outlined in Section 6.3.7 are represented as states. We split each session into two halves to demonstrate what is likely to happen at the start of the session when compared to the end of the session.

From Figure 6.4, we can first observe that groups in the SepNav condition had more search events in the first half of the session (62.90% of the events) than in the second half (39.64% of the events). Consequently, SepNav participants communicated less in the first half (28.82% of the events) than in the second half (40.28% of the events). In contrast, participants in SepNav condition to communicate more uniformly along with the session. Also, once a search event occurs, participants in the SharedNav condition are more likely in the first half of the session to send a message to their collaborators (31.32%) than SepNav participants (17.60%). This suggests that participants in the SharedNav condition alternate more between communicating and **Stage 1** and **Stage 2** of the decision making process. This is an important implication for future research as we can identify key timestamps of



(a) SepNav

(b) SharedNav

Figure 6.4: Markov model constructed based on the transitions between events described in Table 6.2.

Table 6.4: Post-task questions regarding post-purchase satisfaction and system perception. Mean values (± standard deviations) are reported across conditions SepNav and SharedNav. A dagger (†) next to SharedNav values denotes significance from the baseline, SepNav.

Measure			SepNav	SharedNav
Post-purchase	**PP1**	**Purchase decision satisfaction** (1 (not satisfied) - 7 (satisfied))	6.44(±1.16)	6.16(±1.44)
	PP2	**Search space exploration** (1 (not sufficient) - 7 (sufficient))	5.94(±1.27)	5.75(±1.41)
	PP3	**Discussed with others** (1 (not sufficient) - 7 (sufficient))	6.09(±1.28)	6.28(±1.08)
	PP4	**Expressed opinions** (1 (not sufficient) - 7 (sufficient))	6.19(±1.18)	6.38(±0.75)
	PP5.	**Own opinions impacted** (1 (disagree) - 7 (agree))	6.03(±1.23)	6.28(±1.11)
System Perception	**SP1**	**Relevance of search results** (1 (disagree) - 7 (agree))	5.62(±1.41)	5.06(±1.50)†
	SP2	**No system technical issues** (1 (disagree) - 7 (agree))	4.97(±1.99)	3.88(±1.93)†
	SP3	**Task difficulty** (1 (easy) - 7 (difficult))	3.34(±1.60)	2.97(±1.58)
	SP4	**Easiness of synchronisation** (1 (easy) - 7 (difficulty))	5.62(±1.41)	5.06(±1.50)
	SP5	**Awareness of each other** (1 (disagree) - 7 (agree))	5.47(±1.70)	5.16(±1.71)
	SP6	**Easiness of sharing knowledge** (1 (easy) - 7 (difficulty))	4.97(±1.99)	3.88(±1.93)

the session to facilitate the shared navigation among collaborators via chat features.

In terms of review events, SharedNav groups are less likely to continue in **Stage 3** during the first half of the session than SepNav groups. This is expected as group members have to explore more the saved items in order to coordinate and communicate their findings. Surprisingly, during the entire session, less than 0.01% of transition events occurred between search and purchase events, suggesting that in both conditions, there is no direct transition to **Stage 4** from **Stage 1** and **Stage 2**. Additionally, we observe in both conditions that participants triggered search events after a purchase event (14.89% and 13.16%, for SepNav and SharedNav, respectively). Further analysis shows that this transition back to **Stage 1** and **Stage 2** from **Stage 4** is due to two reasons: *(i)* to only conduct more searches and confirm the purchase decision was final (7 groups and 6 groups in the SepNav and SharedNav condition, respectively); or *(ii)* to save and review more products and remake the purchase decision (3 groups and 5 groups in the SepNav and SharedNav condition, respectively).

6.4.3 Post-Purchase and System Perception

Finally, we present the perceived post-purchase satisfaction results in Table 6.4, or **Stage 5** of the decision-making process (row **PP1–5**). As we ran a simulated task, we are unable to capture how participants evaluate their purchase besides qualitative measurements. In this stage, we did not observe statistical differences for the perceived purchase decision satisfaction. This finding means that collaborative navigation did not impact the perceived satisfaction with the collaborative shopping task. On the other hand, in terms of system perception (row **SP1–6**), we observe that participants subjected to `SharedNav` found the search results less relevant than `SepNav` condition participants and that they perceived more technical issues with the system. This can be explained as the collaborative navigation is not something that participants were used to, even though we introduced the system and the features during the training task.

6.5 Limitations

In this section, we shortly describe the limitations of this chapter as our experimental setup had limitations related to: (i) the CSE setting; (ii) participants recruitment; and (iii) the task descriptions that we designed. First, we acknowledge that the CSE setting in this study was limited to groups of two and three members. Second, we recruited participants from a technological university, limiting demographic diversity. Last, we designed the tasks descriptions as realistic as possible. However, we acknowledge that not all participants had previously experienced a collaborative shopping task. Overall, we believe that our contributions in this chapter are still valid and are valuable results as a reference for future work.

6.6 Conclusions

In this chapter, we aim to answer the following research question:

RQ4 How and when do navigational mechanisms impact the group decision making process in collaborative online shopping tasks?

To this end, we investigated how the five stages of group decision (Section 6.2.3) making are impacted by the type of navigational support (separate vs. shared navigation) in a collaborative product search task. We conducted a user study with 64 participants. Participants were split between two systems: one that sought to have individuals search individually, with the other introducing shared navigation between group members. We observed that of the stages of group decision making, **Stages 1–3** were the most impacted by the navigation support. In particular, we found that participants of shared navigation were more effective in their search space exploration, and shared navigation caused less usage of other CSE features, such as viewing recent queries. We observed that shared navigation elicits participants to communicate more consistently during the task, and no significant differences were found across the metrics we considered for **Stage 4** and **Stage 5**.

Implications of the work conducted in this chapter demonstrate that researchers should consider analysing the outcomes of online collaborative shopping tasks using the five-Stage framework by [75], as it provides a more structured manner to understand the complex processes involved. We confirmed that collaborative navigation does aid searchers when collaboratively shopping. We also confirmed that the inclusion of chat tools also aids the process; not being co-located, chat provides an invaluable means of communication between group members. To the best of our knowledge, this study is the first that looks at user behaviour measures across independent and shared navigation systems in e-commerce. Our insights can be useful as they open up a number of possible future research directions, such as:

- Explore which types of products collaborative shopping and navigation support may be beneficial. For example, study the situations where products are of high consideration, which take longer to find or increase the risk of buying them.
- Identifying other domains where collaborative navigation may be suitable. A promising domain is literature search, as researchers conducting literature reviews may benefit from shared navigation to brainstorm and discuss previous literature.
- Investigate how a chatbot facilitator can improve group chat discussions when shared navigation may not be as effective, for instance, in groups larger than three members.

7

Conclusions

In this thesis, we examined the effectiveness of CSE in two parts: (i) the impact of differences in *group configuration* on CSE system effectiveness, and (ii) effectiveness of CSE systems in support of collaborative tasks in the context of search as learning and collaborative online shopping. We presented four user studies with more than 700 participants and described their findings.

In this concluding chapter, we first revisit the research questions we introduced in Chapter 1 and describe our main findings in Section 7.1. Then, in Section 7.3, we discuss future directions, and conclude with final remarks in Section 7.4.

7.1 Main Findings

7.1.1 Effectiveness on Supporting Group Configurations

Within the first part of this thesis, we answered two research questions motivated by the need of investigating understudied group configurations, namely, *group sizes*, and *users' roles*. First, in Chapter 3, we turned our attention to the impact of group size on CSE effectiveness, which has not received much attention in past research. Previous work based on simulations investigated the effect of group size changes [70]. However, simulations are limited in their ability to model the real world. Thus, we conducted an elaborated user study to investigate to what extent the findings of this simulation study hold in a setup with actual users. To this end, we asked the following question:

RQ1 What is the impact of group size on CSE in recall-oriented tasks in terms of retrieval effectiveness and search behaviours?

To answer this question, we designed a crowd-sourcing based user experiment with more than 300 crowd-workers across group sizes of two, four, and six, and three CSE system variants. Our findings show that we did not observe diminishing returns with increasing group sizes; the group recall steadily increases as more collaborators participate in the search. We also did not find larger collaborating groups beneficial at the start of a search session mostly. Instead, the increased recall obtained early in the search session compared to smaller collaborating groups is retained throughout the search session. Lastly, we noted that our results confirmed our intuition that division of labour and sharing of knowledge approaches need to be considered with care as group sizes increase.

Next, in Chapter 4, we addressed the weakness of previous works on providing sufficient evidence to the utility of assigned roles in CSE. The main drawback of these studies was that each one was limited to comparing only two experimental conditions. Thus, in Chapter 4, we set out to settle the following question:

RQ2 What are the benefits of role-based CSE recall-oriented tasks in terms of search retrieval effectiveness, search behaviours, and communication patterns?

To answer this question, we designed a crowd-sourcing based user study across four conditions: (i) individual searchers; (ii) artificial groups formed by aggregating the search logs of pairs of individual searchers; (iii) pairs of collaborating searchers without assigned roles; and (iv) pairs of collaborating searchers with assigned roles. Besides search effectiveness, we investigated the differences in search behaviour and communication patterns between those conditions.

Although our search behaviour and communication pattern analyses exhibit a substantial difference among those groups, we found that search effectiveness does not differ significantly between groups with and without assigned roles. Particularly, we found that our retrieval effectiveness measures (precision and recall) do not differ significantly between groups with and without assigned roles. Artificial groups (i.e., pairs of individual searchers) did not benefit from communication and coordination perform at the same level of recall as collaborative groups. In addition, we found that groups with assigned roles show search behaviours that indicate their compliance with the given role assignment: group members tasked with exploring the search space spent less time reading documents, and issued more queries. In contrast to groups without pre-assigned roles, group members assigned the task of exploring the set of documents pre-selected by another group member spent more time reading documents. Despite the apparent differences in search behaviours, search performance overall does not differ significantly. Lastly, we found that groups with assigned roles communicate less frequently than groups without assigned roles. Again, though, the time saved from fewer exchanges did not turn into an improved search performance. We concluded that role-based CSE does not seem to be beneficial as users alone can successful self-organise the collaborating group to reach their goals.

7.1.2 Supporting Search as Learning and Online Shopping Contexts

We then moved to the second part of this thesis. We focused on two contexts, namely, search as learning, and collaborative online shopping. In Chapter 5, we turned our attention to search as learning and answered the research question:

RQ3 Is (individual and collaborative) search a viable alternative to instructor-designed learning?

To answer this question, we designed a crowd-sourced user study with 151 participants assigned to one of the five different conditions that consisted of instructor-designed learning (short high-quality video lecture) against three forms of search. In particular, our experiment conditions consisted of (i) search-only, (ii) search as a support tool for instructor-designed learning and (iii) collaborative search. We measured their learning gains in a vocabulary learning task, which proved to be sufficiently challenging though testing only lower cognitive skill levels. We found that participants in the instructor-designed learning condition reached up to 24% higher learning gains than participants in the search-only condition. Also, instructor-designed learning supported by search is superior to instructor-designed learning alone, leading to a 41% increase in realised potential learning. At the same time, these increases in learning gain do not translate into a higher confidence in the search process. Lastly, we found that CSE is not competitive as the collaborative cost leaves less time for the retrieval and sensemaking steps of the search process: the learning gains decreased significantly compared to the video conditions. Considering that we used very high-quality lecture videos in comparison to search, we consider these results to indicate that search can be a viable (though worse) alternative to instructor designed learning, especially in situations where no high-quality video material is available.

Our last study was in navigation support in collaborative shopping tasks. In Chapter 6, we addressed the following research question:

RQ4 How and when do navigational mechanisms impact the group decision making process in collaborative product search tasks?

To answer this question, we conducted a user study with 64 participants. Participants were assigned to one of the two experimental conditions: separate or shared navigation. We hypothesised that shared navigation would impact the group decision making process stages by enabling increased co-presence and consequently increased awareness and sharing of knowledge of the CSE process. We found that of the stages of group decision making, the stages of search, review, and evaluate, were the most impacted by the navigation support. In particular, we found that participants of shared navigation were more effective in their search space exploration, and shared navigation caused less usage of other CSE features, such as viewing recent queries. In addition, we observed that shared navigation elicits participants to communicate more consistently during the task, and no significant differences were found across the metrics we considered for the purchase decision and post-purchase decision stages of the group decision making process. We believe that the work done in this chapter demonstrates that considering online collaborative shopping tasks must be evaluated with the group decision making stages in mind, as it provides a more structured manner to understand the complex processes involved. We confirmed that collaborative navigation does help searchers when collaboratively shopping. Lastly, we also confirmed that the inclusion of chat tools also aids the process; not being co-located, chat provides an invaluable means of communication between group members. To the best of our knowledge, this study is the first that looks at user behaviour measures across independent and shared navigation systems in e-commerce.

7.2 Limitations

In this thesis, the user studies we conducted in this thesis had overall limitations related to: (i) the recruitment of the participants for the studies; (ii) user study instructions; and (iii) crowd-sourcing experiments during the Coronavirus pandemic. First, in Chapters 3–5, we recruited participants via a crowd-sourcing platform, where we limited our participants to

English speakers, and we paid a contribution according to the Prolific platform's suggestion. In Chapter 6, our participants consisted of mostly students from tech backgrounds, which also could be limited due to the less diverse background. We acknowledge that the participants could be limited, but we believe it does not affect the relevance of the presented results in those chapters. Second, we designed our studies and gave instructions to the participants for doing tasks that would be as realistic as possible. Although many of these tasks were inspired by related works, we acknowledge that our results could be affected in terms of generalisability. Lastly, we recruited participants for two studies (Chapter 4 and Chapter 6) during a period of the Coronavirus pandemic (Europe lockdown in April 2020 and April 2021). We acknowledge that the increased distractions at home and workplace (e.g., with homeschooling or family care) could have affected the participants' performance during the experiments since most of our participants were located in countries with rigid lockdown measures (most of our participants resided in the UK and The Netherlands).

7.3 Future Directions

In the following, we discuss directions for future work to further expand our work.

7.3.1 Evaluation of Collaborative Search Systems

Evaluation in CSE is a continually discussed and difficult issue among CSE researchers, with two workshops on CSE evaluation having been held in 2015 and 2017 (ECol'2015 [137] and ECol'2017 [13]). The goal of the first workshop in 2015 was to identify issues on evaluating CSE and why CSE evaluation is more challenging than individual interactive search evaluation. Throughout this thesis, we leveraged many aspects from IIR and HCI research fields to help us make our study design decisions. Similarly, previous and concurrent CSE studies also used metrics and data collections common in IIR studies. However, CSE studies have many aspects that differ from IIR studies, such as group configurations and collaborative tasks. In the second workshop in 2017, participants identified the focus of evaluation in CSE should be on the understandability, usability and quality of outcomes, along with knowledge sharing and sensemaking. However, participants also considered these latter evaluation aspects much more challenging to measure and evaluate.

As a result of the discussions of the two workshops, there is space for formalising an evaluation framework for CSE. This framework should take into account the many aspects that differ CSE studies from IIR studies and provide a clear guideline to CSE researchers. Consequently, this could facilitate researchers in making decisions during the study design stage and provide robust experimentation for CSE.

7.3.2 Modelling and Simulating Collaborative Search Behaviour

Modelling user behaviour in IIR has been extensively studied with formal models such as economic models [10]. Search and stop behaviour models have oriented researchers in identifying the costs and benefits of interface features such as a recent query widget. Moreover, such a model has been the base for robust simulation in IIR and provides an alternative and complement to user studies. A future line of work in CSE is to model collaborators' search behaviour. Previous works have used simulation of IIR behaviour to simulate a CSE session. However, those simulations did not consider modelling collaborators' individual behaviour and interactions with other collaborators. Future work should take into account the interaction behaviour of collaborators and model other CSE characteristics such as group sizes and users' roles. This line of work could potentially assist fast prototyping of interface features as well algorithmic mediation approaches.

7.3.3 Collaborative Learning to Rank Models

Collaborative ranking models have been developed in a two steps fashion [134]. For instance, the first step divides a data collection given a query (e.g., with clustering techniques), and the second matches potential documents to each collaborator given their profile (e.g., expertise) [136, 138]. Although these two steps provide a clear modelling of collaborative rankers, another set of collaborative rankers could be developed in one step based on learning to rank models and optimising one objective instead of two.

Collaborative learning to rank models could use as input the user, group, query representations and produce as output the tailored list of ranked documents for the individual user. However, to train learning to rank models, data collections and query-document relevant pairs are required. One challenge that arises is that, in CSE, the query-document relevance score can be different for each user in a group. Here, simulation can help build learning to rank models by providing simulated behaviour data to train the models. Ultimately, learning to rank models can be evaluated via user studies and learn from collaborators' behaviour as they interact with the search results.

7.3.4 Collaborative Conversational Search

Conversation is an important aspect in CSE when collaborators are synchronised in a search session. Throughout, we have recognised the importance of conversation and analysed the conversation behaviour of collaborators. Recently, a lot of effort of IR researchers has gone into conversational search data collections and evaluation methodologies [38]. However, the number of works that focused on the conversation aspect of CSE has been limited. Previous studies focused on comparing voice versus text conversation [55], however, only a few to formalise how collaborators would discuss their information need and divide the task via conversation [132, 133].

Therefore, a promising line of research is to identify attributes of conversation in CSE that can be beneficial to help collaborators, for instance, discuss their information need, divide the task, and share knowledge. Another attribute that would be worth considering is the impact of group sizes in conversational CSE. In particular, adding more users to a group could hinder the group chat from becoming inviable to keep up with their searching and chatting.

Finally, in the opposite direction of this thesis, one can continue the development of CSE by only focusing on supporting CSE via conversational systems as conducted by Avula [8] or with voice-enabled CSE with voice assistants such as Amazon Alexa and Google Home. Here, an open question that is interesting to answer is what type of task conversational CSE can support effectively [61].

7.4 Final Remarks

Examining the effectiveness of CSE systems is a challenging endeavour that involves many research and engineering decisions. As demonstrated throughout the thesis, we have examined CSE effectiveness in several dimensions, focused on designing user research studies that could provide us evidence on the benefits of CSE systems. We have also provided an open source framework, `SearchX`, that we believe will assist the research community. Our investigations considered many aspects of the `SearchX`'s UI such as the collaborative features usage. We believe we have provided a robust CSE system for future researchers to conduct experiments with a focus more on algorithmic mediation approaches. Finally, we believe that this thesis will potentially aid researchers in further exploring many aspects of CSE systems.

Bibliography

[1] James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012. In *ACM SIGIR Forum*, pages 2–32, 2012.

[2] Omar Alonso. The practice of crowdsourcing. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 11(1):1–149, 2019.

[3] Saleema Amershi and Meredith Ringel Morris. Cosearch: a system for co-located collaborative web search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1647–1656, 2008.

[4] Lorin W Anderson, David R Krathwohl, P Airasian, K Cruikshank, R Mayer, P Pintrich, James Raths, and M Wittrock. A taxonomy for learning, teaching and assessing: A revision of bloom's taxonomy. *New York. Longman Publishing. Artz, AF, & Armour-Thomas*, pages 137–175, 2001.

[5] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11):2635–2651, 2016.

[6] Kumaripaba Athukorala, Alan Medlar, Antti Oulasvirta, Giulio Jacucci, and Dorota Glowacka. Beyond relevance: Adapting exploration/exploitation in information retrieval. In *Proceedings of the 21st international conference on intelligent user interfaces*, pages 359–369, 2016.

[7] Mohsen Attaran, Sharmin Attaran, and Diane Kirkland. The need for digital workplace: increasing workforce productivity in the information age. *International Journal of Enterprise Information Systems (IJEIS)*, 15(1):1–23, 2019.

[8] Sandeep Avula. *Characterizing and Understanding User Perception of System Initiative for Conversational Systems to Support Collaborative Search.* PhD thesis, The University of North Carolina at Chapel Hill, 2020.

[9] Sandeep Avula, Gordon Chadwick, Jaime Arguello, and Robert Capra. Searchbots: user engagement with chatbots during collaborative search. In *Proceedings of the 2018 conference on human information interaction & retrieval*, pages 52–61, 2018.

[10] Leif Azzopardi. Modelling interaction with economic models of search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 3–12, 2014.

[11] Leif Azzopardi, Diane Kelly, and Kathy Brennan. How query cost affects search behavior. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 23–32, 2013.

[12] Leif Azzopardi, Jeremy Pickens, Chirag Shah, Laure Soulier, and Lynda Tamine. Second international workshop on the evaluation of collaborative information seeking and retrieval (ecol'17). In *CHIIR '17*, pages 429–431, 2017.

[13] Leif Azzopardi, Jeremy Pickens, Chirag Shah, Laure Soulier, and Lynda Tamine. Report on the second international workshop on the evaluation on collaborative information seeking and retrieval (ecol2017@ chiir). In *ACM SIGIR Forum*, volume 51, pages 122–127. ACM New York, NY, USA, 2018.

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

[15] Ricardo Baeza-Yates and José A Pino. Towards formal evaluation of collaborative work. *Information Research: An International Electronic Journal*, 11(4):n4, 2006.

[16] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern information retrieval: the concepts and technology behind search. *New Jersey, USA: Addi-son-Wesley Professional*, 2011.

[17] Meredith Belbin. Belbin team roles. *Book Belbin Team Roles*, 2004.

[18] Nicholas Belkin, Toine Bogers, Jaap Kamps, Diane Kelly, Marijn Koolen, and Emine Yilmaz. Second workshop on supporting complex search tasks. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 433–435, 2017.

[19] Nicholas J Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1):133–143, 1980.

[20] Sumit Bhatia, Debapriyo Majumdar, and Prasenjit Mitra. Query suggestions in the absence of query logs. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 795–804, 2011.

[21] Keping Bi, Qingyao Ai, Yongfeng Zhang, and W Bruce Croft. Conversational product search based on negative feedback. In *Proceedings of the 28th acm international conference on information and knowledge management*, pages 359–368, 2019.

[22] J Patrick Biddix, Chung Joo Chung, and Han Woo Park. Convenience or credibility? a study of college student online research behaviors. *The Internet and Higher Education*, pages 175–182, 2011.

[23] Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. Towards a probabilistic model for supporting collaborative information access. *Information Retrieval Journal*, 19 (5):487–509, 2016.

[24] Susan E Brennan, Xin Chen, Christopher A Dickinson, Mark B Neider, and Gregory J Zelinsky. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465–1477, 2008.

[25] Marc Bron, Jasmijn Van Gorp, Frank Nack, Lotte Belice Baltussen, and Maarten de Rijke. Aggregated search interface preferences in multi-session search tasks. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 123–132, 2013.

[26] Martynas Buivys and Leif Azzopardi. Pienapple search: an integrated search interface to support finding, refinding and sharing. *Proceedings of the Association for Information Science and Technology*, (1):1–5, 2016.

[27] Katriina Byström and Preben Hansen. Conceptual framework for tasks in information studies. *Journal of the American Society for Information science and Technology*, 56(10):1050–1061, 2005.

[28] Arthur Câmara, Nirmal Roy, David Maxwell, and Claudia Hauff. Searching to learn with instructional scaffolding. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 209–218, 2021.

[29] Robert Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 951–960, 2010.

[30] Robert Capra, Annie T Chen, Katie Hawthorne, Jaime Arguello, Lee Shaw, and Gary Marchionini. Design and evaluation of a system to support collaborative search. *Proceedings of the American Society for Information Science and Technology*, 49(1): 1–10, 2012.

[31] Robert Capra, Annie T Chen, Katie Hawthorne, Jaime Arguello, Lee Shaw, and Gary Marchionini. Design and evaluation of a system to support collaborative search. *Proceedings of the American Society for Information Science and Technology*, 49(1): 1–10, 2012.

[32] Robert Capra, Annie T Chen, Evonne McArthur, and Natalie Davis. Searcher actions and strategies in asynchronous collaborative search. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–10, 2013.

[33] Yihong Cheng, Yanzhen Yue, Zhenhui Jack Jiang, and Hyung Jin Kim. The effects of navigation support and group structure on collaborative online shopping. In *International Conference on Online Communities and Social Computing*, pages 250–259. Springer, 2013.

[34] Amelia W Cole. Engagement in metacognition within self-determined learning. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 495–498, 2020.

[35] Michael J Cole, Jacek Gwizdka, Chang Liu, Nicholas J Belkin, and Xiangmin Zhang. Inferring user knowledge level from eye movement patterns. *Information Processing & Management*, 49(5):1075–1091, 2013.

[36] Kevyn Collins-Thompson, Soo Young Rieh, Carl C Haynes, and Rohail Syed. Assessing learning outcomes in web search: A comparison of tasks and query strategies. In *Proceedings of the 2016 ACM on conference on human information interaction and retrieval*, pages 163–172, 2016.

[37] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. Search as Learning (Dagstuhl Seminar 17092). *Dagstuhl Reports*, pages 135–162, 2017.

[38] J Shane Culpepper, Fernando Diaz, and Mark D Smucker. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (swirl 2018). In *ACM SIGIR Forum*, volume 52, pages 34–90. ACM New York, NY, USA, 2018.

[39] Edward Cutrell and Zhiwei Guan. What are you looking for? an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416, 2007.

[40] Edgar Dale. Vocabulary measurement: Techniques and major findings. *Elementary English*, pages 895–948, 1965.

[41] Abdigani Diriye and Gene Golovchinsky. Querium: a session-based collaborative search system. In *European Conference on Information Retrieval*, pages 583–584. Springer, 2012.

[42] Robert Dorfman. A formula for the gini coefficient. *The review of economics and statistics*, pages 146–149, 1979.

[43] Katherine A Dougherty Stahl and Marco A Bravo. Contemporary classroom vocabulary assessment for content areas. *The Reading Teacher*, pages 566–578, 2010.

[44] Ashlee Edwards and Diane Kelly. How does interest in a work task impact search behavior and engagement? In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, pages 249–252, 2016.

[45] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 223–232, 2014.

[46] Colum Foley and Alan F Smeaton. Synchronous collaborative information retrieval: Techniques and evaluation. In *European Conference on Information Retrieval*, pages 42–53. Springer, 2009.

[47] Colum Foley and Alan F Smeaton. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *Information processing & management*, 46(6):762–772, 2010.

[48] Jonathan Foster. Collaborative information seeking and retrieval. *Annual Review of Information Science and Technology*, 40(1):329–356, December 2006. ISSN 0066-4200.

[49] Yanjun Gao, Madhu Reddy, and Bernard Jansen. Shopwithme!: Collaborative information searching and shopping for online retail. In *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2017.

[50] Urs Gasser, Sandra Cortesi, Momin M Malik, and Ashley Lee. Youth and digital media: From credibility to information quality. *Berkman Center Research Publication*, 2012.

[51] Gene Golovchinsky, John Adcock, Jeremy Pickens, Pernilla Qvarfordt, and Maribeth Back. Cerchiamo: a collaborative exploratory search tool. *Proceedings of Computer Supported Cooperative Work (CSCW)*, pages 8–12, 2008.

[52] Gene Golovchinsky, Jeremy Pickens, and Maribeth Back. A taxonomy of collaboration in online information seeking. *JCDL Workshop on Collaborative Information Retrieval*, 2009.

[53] Gene Golovchinsky, Pernilla Qvarfordt, and Jeremy Pickens. Collaborative information seeking. *Computer*, (3):47–51, 2009.

[54] Gene Golovchinsky, Abdigani Diriye, and Tony Dunnigan. The future is in the past: designing for exploratory search. In *Proceedings of the 4th information interaction in context symposium*, pages 52–61, 2012.

[55] Roberto González-Ibáñez, Muge Haseki, and Chirag Shah. Let's search together, but not too close! an analysis of communication and performance in collaborative information seeking. *Information Processing & Management*, 49(5):1165–1179, 2013.

[56] Philip J Guo, Juho Kim, and Rob Rubin. How video production affects student engagement: An empirical study of mooc videos. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 41–50, 2014.

[57] Yangyang Guo, Zhiyong Cheng, Liqiang Nie, Yinglong Wang, Jun Ma, and Mohan Kankanhalli. Attentive long short-term preference modeling for personalized product search. *ACM Transactions on Information Systems (TOIS)*, 37(2):1–27, 2019.

[58] Preben Hansen and Kalervo Järvelin. Collaborative information retrieval in an information-intensive domain. *Information Processing & Management*, 41(5):1101–1119, 2005.

[59] Ahmed Hassan Awadallah, Ryen W White, Patrick Pantel, Susan T Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 829–838, 2014.

[60] Gerald Häubl and Valerie Trifts. Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science*, 19(1):4–21, 2000.

[61] Marti A Hearst. What's missing from collaborative search? *Computer*, 47(3):58–61, 2014.

[62] Jason Hill and Carl Gutwin. The maui toolkit: Groupware widgets for group awareness. *Computer Supported Cooperative Work (CSCW)*, 13(5):539–571, 2004.

[63] Nyi Nyi Htun, Martin Halvey, and Lynne Baillie. Towards quantifying the impact of non-uniform information access in collaborative information retrieval. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 843–846, 2015.

[64] Nyi Nyi Htun, Martin Halvey, and Lynne Baillie. How can we better support users with non-uniform information access in collaborative information retrieval? In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 235–244, 2017.

[65] Shan Huang, Izak Benbasat, and Andrew Burton-Jones. The role of product recommendation agents in collaborative online shopping. In *ICIS*, 2011.

[66] Kosetsu Ikeda, Atsuyuki Morishima, Habibur Rahman, Senjuti Basu Roy, Saravanan Thirumuruganathan, Sihem Amer-Yahia, and Gautam Das. Collaborative crowdsourcing with crowd4u. *Proceedings of the VLDB Endowment*, 9(13):1497–1500, 2016.

[67] Behnam Izadi, Linying Dong, and Mohammad Rahim Esfidani. Stay home and shop together. *Journal of Electronic Commerce Research*, 22(1):59, 2021.

[68] Rodriguez Perez Jesus, Teerapong Leelanupab, and Joemon M Jose. Cofox: A synchronous collaborative browser. In *Asia Information Retrieval Symposium*, pages 262–274. Springer, 2012.

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

[70] Hideo Joho, David Hannah, and Joemon M Jose. Revisiting ir techniques for collaborative search strategies. In *European Conference on Information Retrieval*, pages 66–77. Springer, 2009.

[71] Epaminondas Kapetanios. Quo vadis computer science: From turing to personal computer, personal content and collective intelligence. *Data & Knowledge Engineering*, 67(2):286–292, 2008.

[72] Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and trends in Information Retrieval*, 3(1—2):1–224, 2009.

[73] Ryan Kelly and Stephen J Payne. 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*, pages 807–819, 2014.

[74] Hongki Kim, Kil-Soo Suh, and Un-Kon Lee. Effects of collaborative online shopping on shopping experience through social and relational perspectives. *Information & Management*, 50(4):169–180, 2013.

[75] Young Ae Kim and Jaideep Srivastava. Impact of social influence in e-commerce decision making. In *Proceedings of the ninth international conference on Electronic commerce*, pages 293–302, 2007.

[76] Simon Knight, Bart Rienties, Karen Littleton, Dirk Tempelaar, Matthew Mitsui, and Chirag Shah. The orchestration of a collaborative information seeking learning task. *Information Retrieval Journal*, 20(5):480–505, 2017.

[77] Michail D Kozlov and Cornelia S Große. Online collaborative learning in dyads: Effects of knowledge distribution and awareness. *Computers in Human Behavior*, 59:389–401, 2016.

[78] Udo Kruschwitz and Charlie Hull. Searching the enterprise. *Foundations and Trends in Information Retrieval*, 11:1–142, 2017.

[79] Bill Kules and Robert Capra. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. *Journal of the American Society for Information Science and Technology*, 63(1):114–138, 2012.

[80] Bill Kules and Ben Shneiderman. Users can change their web search tactics: Design guidelines for categorized overviews. *Inf. Process. Manage.*, 44(2), March 2008.

[81] Anand Kulkarni, Matthew Can, and Björn Hartmann. Collaboratively crowdsourcing workflows with turkomatic. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 1003–1012, 2012.

[82] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. Right way, right time: Towards a better comprehension of young students' needs when looking for relevant search results. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, pages 256–261, 2021.

[83] Victor Lavrenko and W Bruce Croft. Relevance-based language models. In *ACM SIGIR Forum*, volume 51, pages 260–267. ACM New York, NY, USA, 2017.

[84] Teerapong Leelanupab, Hannarin Kruajirayu, and Nont Kanungsukkasem. Snapboard: A shared space of visual snippets-a study in individual and asynchronous collaborative web search. In *AIRS*, pages 161–173. Springer, 2015.

[85] Olivier Liechti. Awareness and the www: An overview. *SIGGROUP Bulletin*, 21(3): 3–12, December 2000. ISSN 2372-7403.

[86] Craig Macdonald, Richard McCreadie, Rodrygo L. T. Santos, and Iadh Ounis. From puppy to maturity: Experiences in developing Terrier. In *Proceedings of the SIGIR 2012 Workshop on Open Source Information Retrieval*, pages 60–63, Portland, OR, USA, 2012.

[87] Gurmeet Singh Manku, Arvind Jain, and Anish Das Sarma. Detecting near-duplicates for web crawling. In *Proceedings of the 16th international conference on World Wide Web*, pages 141–150, 2007.

[88] Gary Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[89] David Maxwell and Claudia Hauff. Logui: Contemporary logging infrastructure for web-based experiments. In *European Conference on Information Retrieval*, pages 525–530, 2021.

[90] David Maxwell, Leif Azzopardi, and Yashar Moshfeghi. The impact of result diversification on search behaviour and performance. *Information Retrieval Journal*, 22 (5):422–446, 2019.

[91] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015.

[92] Bhaskar Mitra and Nick Craswell. Query auto-completion for rare prefixes. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1755–1758, 2015.

[93] Abu Shamim Mohammad Arif, Jia Tina Du, and Ivan Lee. Understanding tourists' collaborative information retrieval behavior to inform design. *Journal of the Association for Information Science and Technology*, 66(11):2285–2303, 2015.

[94] Felipe Moraes and Claudia Hauff. node-indri: Moving the indri toolkit to the modern web stack. In *European Conference on Information Retrieval*, pages 241–245. Springer, 2019.

[95] Felipe Moraes, Sindunuraga Rikarno Putra, and Claudia Hauff. Contrasting search as a learning activity with instructor-designed learning. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 167–176, 2018.

[96] Felipe Moraes, Kilian Grashoff, and Claudia Hauff. On the impact of group size on collaborative search effectiveness. *Information Retrieval Journal*, 22(5):476–498, 2019.

[97] Felipe Moraes, Jie Yang, Rongting Zhang, and Vanessa Murdock. The role of attributes in product quality comparisons. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 253–262, 2020.

[98] Felipe Moraes, David Maxwell, and Claudia Hauff. Exploring collaborative navigation support in collaborative product search. In *Proceedings of the SIGIR Workshop on eCommerce*, 2021.

[99] Meredith Ringel Morris. Collaborating alone and together: Investigating persistent and multi-user web search activities. In *Proc. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'07)*, pages 23–27, 2007.

[100] Meredith Ringel Morris. A survey of collaborative web search practices. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1657–1660, 2008.

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

[102] Meredith Ringel Morris and Eric Horvitz. Searchtogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 3–12, 2007.

[103] Meredith Ringel Morris, Jaime Teevan, and Steve Bush. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting. In *CSCW '08*, pages 481–484, 2008.

[104] Meredith Ringel Morris, Jarrod Lombardo, and Daniel Wigdor. Wesearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 401–410, 2010.

[105] Tatsuya Nakamura, Tomu Tominaga, Miki Watanabe, Nattapong Thammasan, Kenji Urai, Yutaka Nakamura, Kazufumi Hosoda, Takahiro Hara, and Yoshinori Hijikata. Investigation on dynamics of group decision making with collaborative web search. In *Proceedings of the International Conference on Web Intelligence*, pages 434–441, 2017.

[106] David Nicholas, Ian Rowlands, David Clark, and Peter Williams. Google generation ii: web behaviour experiments with the bbc. In *Aslib proceedings*. Emerald Group Publishing Limited, 2011.

[107] Christian Otto, Ran Yu, Georg Pardi, Johannes von Hoyer, Markus Rokicki, Anett Hoppe, Peter Holtz, Yvonne Kammerer, Stefan Dietze, and Ralph Ewerth. Predicting knowledge gain during web search based on multimedia resource consumption. In *International Conference on Artificial Intelligence in Education*, pages 318–330. Springer, 2021.

[108] Iadh Ounis, Giambattista Amati, Vassilis Plachouras, Ben He, Craig Macdonald, and Christina Lioma. Terrier: A high performance and scalable information retrieval platform. In *SIGR workshop on open source information retrieval*, 2006.

[109] Sharoda A Paul and Meredith Ringel Morris. Cosense: enhancing sensemaking for collaborative web search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1771–1780, 2009.

[110] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. Beyond the turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70:153–163, 2017.

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

[112] Jeremy Pickens, Gene Golovchinsky, and Meredith Ringel Morris. Proceedings of 1st international workshop on collaborative information seeking. *CoRR*, abs/0908.0583, 2009.

[113] Steven Poltrock, Jonathan Grudin, Susan Dumais, Raya Fidel, Harry Bruce, and Annelise Mark Pejtersen. Information seeking and sharing in design teams. In *2003 International ACM SIGGROUP Conference on Supporting Group Work*, GROUP '03, pages 239–247. ACM, 2003.

[114] Sindunuraga Rikarno Putra, Kilian Grashoff, Felipe Moraes, and Claudia Hauff. On the development of a collaborative search system. In *Design of Experimental Search & Information REtrieval Systems*, pages 76–82, 2018.

[115] Sindunuraga Rikarno Putra, Felipe Moraes, and Claudia Hauff. Searchx: Empowering collaborative search research. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1265–1268. Association for Computing Machinery, 2018.

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

[117] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 75–85, 2014.

[118] Ian Rowlands, David Nicholas, Peter Williams, Paul Huntington, Maggie Fieldhouse, Barrie Gunter, Richard Withey, Hamid R Jamali, Tom Dobrowolski, and Carol Tenopir. The google generation: the information behaviour of the researcher of the future. In *Aslib proceedings*. Emerald Group Publishing Limited, 2008.

[119] Nirmal Roy, Felipe Moraes, and Claudia Hauff. Exploring users' learning gains within search sessions. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 432–436, 2020.

[120] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. How do active reading strategies affect learning outcomes in web search? In *European Conference on Information Retrieval*, pages 368–375, 2021.

[121] Nirmal Roy, Manuel Valle Torre, Ujwal Gadiraju, David Maxwell, and Claudia Hauff. Note the highlight: Incorporating active reading tools in a search as learning environment. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 229–238, 2021.

[122] Tuukka Ruotsalo, Jaakko Peltonen, Manuel JA Eugster, Dorota Głowacka, Aki Reijonen, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. Scinet: interactive intent modeling for information discovery. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1043–1044, 2015.

[123] Sara Salimzadeh, David Maxwell, and Claudia Hauff. The Impact of Entity Cards on Learning-Oriented Search Tasks. In *Proceedings of the 7th ACM SIGIR International Conference on the Theory of Information Retrieval*, 2021.

[124] Chirag Shah. Collaborative information seeking: A literature review. In *Advances in librarianship*, pages 3–33. Emerald Group Publishing Limited, 2010.

[125] Chirag Shah. *Collaborative information seeking: The art and science of making the whole greater than the sum of all*, volume 34. Springer Science & Business Media, 2012.

[126] Chirag Shah and Roberto González-Ibáñez. Evaluating the synergic effect of collaboration in information seeking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 913–922, 2011.

[127] Chirag Shah and Gary Marchionini. Awareness in collaborative information seeking. *Journal of the American Society for Information Science and Technology*, 61(10): 1970–1986, 2010.

[128] Chirag Shah, Gary Marchionini, and Diane Kelly. Learning design principles for a collaborative information seeking system. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3419–3424, 2009.

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

[130] Chirag Shah, Chathra Hendahewa, and Roberto González-Ibáñez. Two's company, but three's no crowd: Evaluating exploratory web search for individuals and teams. *Aslib Journal of Information Management*, 2015.

[131] Chirag Shah, Chathra Hendahewa, and Roberto González-Ibáñez. Two's not always company: collaborative information seeking across task types. *Aslib Journal of Information Management*, 2017.

[132] Sosuke Shiga, Hideo Joho, Roi Blanco, Johanne R Trippas, and Mark Sanderson. Modelling information needs in collaborative search conversations. In *Proceedings*

of the 40th international acm sigir conference on research and development in information retrieval, pages 715–724, 2017.

[133] Keng Siau, Fiona Nah, and Sha H. Efficacy of communication support in collaborative online shopping: The moderating effect of task types. volume 2, 08 2013.

[134] Laure Soulier and Lynda Tamine. On the collaboration support in information retrieval. *ACM Computing Surveys*, 2017.

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

[136] Laure Soulier, Lynda Tamine, and Wahiba Bahsoun. On domain expertise-based roles in collaborative information retrieval. *Information Processing & Management*, 50(5):752–774, 2014.

[137] Laure Soulier, Lynda Tamine, Tetsuya Sakai, Leif Azzopardi, and Jeremy Pickens. Report on the first international workshop on the evaluation on collaborative information seeking and retrieval (ecol'2015). In *ACM SIGIR Forum*, volume 50, pages 42–48. ACM New York, NY, USA, 2016.

[138] Laure Soulier, Lynda Tamine, and Chirag Shah. Minerank: Leveraging users' latent roles for unsupervised collaborative information retrieval. *Information Processing & Management*, 52(6):1122–1141, 2016.

[139] Trevor Strohman, Donald Metzler, Howard Turtle, and W Bruce Croft. Indri: A language model-based search engine for complex queries. In *Proceedings of the international conference on intelligent analysis*, volume 2, pages 2–6, 2005.

[140] Thanaporn Sundaravej, Dinesh Mirchandani, and Albert Lederer. Synchronous collaboration technology use in teamwork. In *2015 48th Hawaii International Conference on System Sciences*, pages 216–225. IEEE, 2015.

[141] Rohail Syed and Kevyn Collins-Thompson. Optimizing search results for human learning goals. *Information Retrieval Journal*, 20(5):506–523, 2017.

[142] Rohail Syed and Kevyn Collins-Thompson. Retrieval algorithms optimized for human learning. In *Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval*, pages 555–564, 2017.

[143] Rohail Syed and Kevyn Collins-Thompson. Exploring document retrieval features associated with improved short-and long-term vocabulary learning outcomes. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 191–200, 2018.

[144] Lynda Tamine and Laure Soulier. Understanding the impact of the role factor in collaborative information retrieval. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 43–52, 2015.

[145] Lynda Tamine and Laure Soulier. Collaborative information retrieval: concepts, models and evaluation. In *European Conference on Information Retrieval*, pages 885–888. Springer, 2016.

[146] Arthur Taylor. A study of the information search behaviour of the millennial generation. *Information Research: An International Electronic Journal*, 2012.

[147] Kimberly Tee, Saul Greenberg, and Carl Gutwin. Providing artifact awareness to a distributed group through screen sharing. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 99–108, 2006.

[148] Jaime Teevan, Kevyn Collins-Thompson, Ryen W White, Susan T Dumais, and Yubin Kim. Slow search: Information retrieval without time constraints. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, pages 1–10, 2013.

[149] Stefan Tilkov and Steve Vinoski. Node. js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83, 2010.

[150] Manos Tsagkias, Tracy Holloway King, Surya Kallumadi, Vanessa Murdock, and Maarten de Rijke. Challenges and research opportunities in ecommerce search and recommendations. In *ACM SIGIR Forum*, volume 54, 2020.

[151] Philip Tuddenham, Ian Davies, and Peter Robinson. Websurface: an interface for co-located collaborative information gathering. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 181–188, 2009.

[152] Melissa A Valentine, Daniela Retelny, Alexandra To, Negar Rahmati, Tulsee Doshi, and Michael S Bernstein. Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 3523–3537, 2017.

[153] Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. Learning latent vector spaces for product search. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 165–174, 2016.

[154] Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. Mix'n match: Integrating text matching and product substitutability within product search. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 1373–1382, 2018.

[155] Ellen M Voorhees. The trec 2005 robust track. In *ACM SIGIR Forum*, volume 40, pages 41–48. ACM, 2006.

[156] Jie Wei, Stefan Seedorf, Paul Benjamin Lowry, Christian Thum, and Thimo Schulze. How increased social presence through co-browsing influences user engagement in collaborative online shopping. *Electronic Commerce Research and Applications*, 24: 84–99, 2017.

[157] Marjorie Wesche and T Sima Paribakht. Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, pages 13–40, 1996.

[158] Ryen W White. *Interactions with search systems.* Cambridge University Press, 2016.

[159] Ryen W White and Resa A Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, pages 1–98, 2009.

[160] Ryen W White, Gheorghe Muresan, and Gary Marchionini. Report on acm sigir 2006 workshop on evaluating exploratory search systems. In *ACM SIGIR Forum*, pages 52–60, 2006.

[161] Ryen W White, Susan T Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*, pages 132–141, 2009.

[162] Mathew J Wilson and Max L Wilson. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the American Society for Information Science and Technology*, 64(2):291–306, 2013.

[163] Luyan Xu, Xuan Zhou, and Ujwal Gadiraju. How does team composition affect knowledge gain of users in collaborative web search? In *Proceedings of the 31st ACM Conference on Hypertext and Social Media*, pages 91–100, 2020.

[164] Peilin Yang, Hui Fang, and Jimmy Lin. Anserini: Enabling the use of lucene for information retrieval research. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1253–1256, 2017.

[165] Ke-Thia Yao, Robert Neches, In-Young Ko, Ragy Eleish, and Sameer Abhinkar. Synchronous and asynchronous collaborative information space analysis tools. In *1999 International Workshops on Parallel Processing*, pages 74–79. IEEE, 1999.

[166] Yanzhen Yue, Xiaojuan Ma, and Zhenhui Jiang. Share your view: impact of co-navigation support and status composition in collaborative online shopping. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3299–3308, 2014.

[167] Zhen Yue, Shuguang Han, and Daqing He. Search tactics in collaborative exploratory web search. *Proceedings of HCIR 2012*, 2012.

[168] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems (TOIS)*, 22(2):179–214, 2004.

[169] Xiangmin Zhang, Michael Cole, and Nicholas Belkin. Predicting users' domain knowledge from search behaviors. In *Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval*, pages 1225–1226, 2011.

[170] Lei Zhu, Izak Benbasat, and Zhenhui Jiang. Let's shop online together: An empirical investigation of collaborative online shopping support. *Information Systems Research*, 21(4):872–891, 2010.

[171] Jie Zou and Evangelos Kanoulas. Learning to ask: Question-based sequential bayesian product search. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 369–378, 2019.

List of Figures

List of Tables

Summary

Although searching is often seen as a solitary activity, searching in collaboration with others is deemed useful or necessary in many complex situations such as: travel planning; online shopping; looking for health related information; planning birthday parties; working on a group project; or finding a house to buy. Researchers have found that complex search tasks can be executed more effectively and efficiently, achieve higher material coverage, and enable higher knowledge gains in an explicit collaborative setting than if conducted in isolation. However, even though researchers have carefully designed several Collaborative Search (CSE) user studies, there is still conflicting evidence or a lack of evidence on the effectiveness of CSE systems. Thus, in this thesis, we focus on examining the effectiveness of CSE systems in two parts.

In the first part, we shed light on the effectiveness of CSE to support two group configurations, namely *group sizes* and *users' roles*.

Past collaborative search studies had a strong focus on groups of two or three collaborators, thus naturally limiting the number of experimental conditions that could increase quickly. Therefore, there is a lack of evidence suggesting the extent to which a CSE system can support group sizes beyond these commonly investigated group sizes. Thus, in Chapter 3, we study CSE system effectiveness with group size as the primary dependent variable. Here, we vary group sizes from two to six collaborators, with six as our upper bound due to limitations on our available resources.

In Chapter 4, we focus on roles in CSE. Roles can determine how a group splits up the search task, and determines each group member's function (e.g., one group member is responsible for finding documents and reading and evaluating them, with a further member responsible for in-depth reading and evaluating of the aforementioned documents). In particular, when the CSE system assigns a role to each group member, researchers have hypothesised that a group may reduce the time spent communicating and coordinating the task, and make the search process more efficient and successful than groups without role assignment. However, past user studies have provided contradicting evidence as to the utility of assigned roles in CSE. Thus, in Chapter 4, we provide more evidence to settle the question of the effectiveness of CSE systems when used by groups with pre-assigned roles versus groups without pre-assigned roles.

In the second part of this thesis, we make our group configurations constant, particularly, group sizes are set to up to three people, and group members receive the same role. We then turn to a different perspective and focus on examining the effectiveness in two contexts: Search as Learning (SAL) and collaborative online shopping.

Search activities for human learning involve multiple iterations that require cognitive processing and interpretation, often requiring the searcher to spend time scanning/viewing, comparing, and evaluating information. However, web search engines are not built to support users in the search tasks often required in learning situations. When people use search as a learning activity, it can be an individual activity or a collaborative activity (e.g.,

group projects). Hence, in Chapter 5, we tackle the challenge of identifying the impact of web search engines on the (single-search or collaborative search) users' ability to learn compared to learning acquired via high-quality learning materials as a baseline.

In Chapter 6, we look at a further context: collaborative online shopping. In collaborative online shopping, a group of people come together to make a decision to purchase a product that meets the various group members' requirements and opinions. While shopping together, search is an important part of the task in order to search for products in a catalogue that is available in an e-commerce website. One important aspect of collaborative shopping is supporting awareness and sharing of knowledge as it can enable a sense of co-presence, which helps groups make a decision that satisfies each group member's requirements and wishes. As search is a significant part of a collaborative online shopping experience, CSE systems are suitable for executing such tasks. However, there is insufficient evidence of how well can CSE systems support a group of users to search for online products together and make a group decision. Hence, in Chapter 6, we explore the effects of increased awareness and sharing of knowledge (co-presence) using a CSE system in collaborative shopping on the group decision making process.

Samenvatting

Hoewel zoeken vaak als een individuele activiteit wordt gezien, wordt zoeken in samenwerking met anderen nuttig of noodzakelijk geacht in veel complexe situaties, zoals: plannen van een reis; online winkelen; zoeken naar gezondheidsgerelateerde informatie; plannen van een verjaardagsfeestje; werken aan een groepsproject; of vinden van een huis om te kopen. Onderzoekers hebben ontdekt dat complexe zoektaken effectiever en efficiënter kunnen worden uitgevoerd, een grotere materiaal-dekking bereiken en meer kennis opleveren in een expliciete omgeving van samenwerking dan wanneer de zoektaken afzonderlijk worden uitgevoerd. Hoewel onderzoekers zorgvuldig verschillende gebruikersonderzoeken voor collaboratieve zoekacties (CSE) hebben opgezet, is er nog steeds tegenstrijdig bewijs of een gebrek aan bewijs over de effectiviteit van CSE-systemen. Daarom richten we ons in dit proefschrift op het onderzoeken van de effectiviteit van CSE-systemen in twee delen.

In het eerste deel werpen we licht op de effectiviteit van CSE om twee groepen configuraties achtereenvolgens te ondersteunen, namelijk *groepsgroottes* en *gebruikersrollen.*

Eerdere onderzoeken naar gezamenlijk zoeken waren sterk gericht op groepen van twee of drie medewerkers, waardoor het aantal experimentele omstandigheden dat snel zou kunnen toenemen, natuurlijk werd beperkt. Daarom is er een gebrek aan bewijs dat suggereert in hoeverre een CSE-systeem groepsgroottes kan ondersteunen die verder gaan dan deze algemeen onderzochte groepsgroottes. Daarom bestuderen we in Hoofdstuk 3 de effectiviteit van het CSE-systeem met groepsgrootte als de primaire afhankelijke variabele. Hier variëren we de groepsgrootte van twee tot zes medewerkers, met zes als onze bovengrens vanwege beperkingen op onze beschikbare middelen.

In Hoofdstuk 4 concentreren we ons op rollen in CSE. Rollen kunnen bepalen hoe een groep de zoektaak opsplitst en zij bepalen de functie van elk groepslid (bijv. één groepslid is verantwoordelijk voor het vinden van documenten en het lezen en evalueren ervan, terwijl een ander lid verantwoordelijk is voor het diepgaand lezen en evalueren van de bovengenoemde documenten). In het bijzonder, wanneer het CSE-systeem een rol toewijst aan elk groepslid, hebben onderzoekers de hypothese aangenomen dat een groep de tijd die besteed wordt aan het communiceren en coördineren van de taak kan verminderen en het zoekproces efficiënter en succesvoller kan maken dan groepen zonder roltoewijzing. Eerdere gebruikersonderzoeken hebben echter tegenstrijdig bewijs opgeleverd over het nut van toegewezen rollen in CSE. In Hoofdstuk 4 geven we dus meer bewijs om de kwestie van de effectiviteit van CSE-systemen op te lossen wanneer ze worden gebruikt door groepen met vooraf toegewezen rollen versus groepen zonder vooraf toegewezen rollen.

In het tweede deel van dit proefschrift maken we onze groepsconfiguraties constant, met name groepsgroottes zijn ingesteld op maximaal drie personen en groepsleden krijgen dezelfde rol. Vervolgens richten we ons op een ander perspectief en richten we ons op het onderzoeken van de effectiviteit in twee contexten: Zoeken als Leren (SAL) en samen online winkelen.

Zoekactiviteiten voor menselijk leren omvatten meerdere iteraties die cognitieve verwerking en interpretatie vereisen, waarbij de zoeker vaak tijd moet besteden aan het scannen/bekijken, vergelijken en evalueren van informatie. Web-zoekmachines zijn echter niet gebouwd om gebruikers te ondersteunen bij de zoektaken die vaak nodig zijn in leersituaties. Wanneer mensen zoeken gebruiken als leeractiviteit, kan dit een individuele activiteit of een samenwerkingsactiviteit zijn (bijvoorbeeld groepsprojecten). Daarom pakken we in Hoofdstuk 5 de uitdaging aan om de impact van web-zoekmachines op het vermogen van gebruikers om te leren (single-search of collaboratief zoeken) te identificeren in vergelijking met een baseline van leren verkregen via hoogwaardig lesmateriaal.

In Hoofdstuk 6 kijken we naar een andere context: collaboratief online winkelen. Bij collaboratief online winkelen komt een groep mensen samen om een beslissing te nemen om een product te kopen dat voldoet aan de eisen en meningen van de verschillende groepsleden. Tijdens het samen winkelen is zoeken een belangrijk onderdeel van de taak om producten te zoeken in een catalogus die beschikbaar is op een e-commerce website. Een belangrijk aspect van collaboratief winkelen is het ondersteunen van het bewustzijn en het delen van kennis, omdat het een gevoel van gezamenlijke aanwezigheid mogelijk kan maken, waardoor groepen een beslissing kunnen nemen die voldoet aan de eisen en wensen van elk groepslid. Aangezien zoeken een belangrijk onderdeel is van een gezamenlijke online winkelervaring, zijn CSE-systemen geschikt voor het uitvoeren van dergelijke taken. Er is echter onvoldoende bewijs van hoe goed CSE-systemen een groep gebruikers kunnen ondersteunen om samen naar online producten te zoeken en een groepsbeslissing te nemen. Daarom onderzoeken we in Hoofdstuk 6 de effecten van een groter bewustzijn en het delen van kennis (co-aanwezigheid) met behulp van een CSE-systeem in collaboratief winkelen op het beslissingsproces van de groep.

Curriculum Vitæ

Felipe Moraes Gomes was born in Divinópolis, Minas Gerais, Brazil on July 1, 1992. He received his masters degree from the Federal University of Minas Gerais, Brazil. His dissertation focused on the effectiveness of dynamic search systems. During his masters, he was a visiting researcher at New York University, USA, where he worked on understanding document representations for document exploration and visualisation. Prior to that, he received his bachelor's degree from the Federal University of Minas Gerais, where he was an undergraduate researcher for three years. During his bachelor's degree, he was a research intern at the University of Quebec at Montreal, Canada. After, he started work as a software engineer intern at Zunnit Technologies, Brazil, where he worked on applied Machine Learning in the news domain.

From May 2017 to November 2021, Felipe was a PhD student in the Web Information Systems group at the Delft University of Technology, supervised by Claudia Hauff. His PhD focused on examining the effectiveness of collaborative search systems by analysing group configurations and application of collaborative search to the search as learning and e-commerce contexts using SearchX, an open-source collaborative search framework that Felipe was the lead contributor. During his PhD, he interned as an applied scientist at Amazon (Seattle, USA), where he conducted research on the role of product attributes in product quality comparisons. Felipe's research has been published in leading conferences and journals on related domains (e.g., ACM SIGIR, ACM CHIIR, BCS-IRSG ECIR, ACM CIKM, ACM ICTIR, and the Information Retrieval Journal). He received the Best Demo Paper Award from ECIR 2019. He also has served as a program committee member and reviewer for several conferences, including ACM SIGIR, BCS-IRSG ECIR, WebConf, EMNLP, ACL, WSDM, CIKM, RecSys. His research interests encompass large-scale web search and recommendation, applied machine learning, and natural language understanding. Felipe is currently a data scientist at Booking.com in Amsterdam, The Netherlands.

Publications

1. Felipe Moraes, David Maxwell and Claudia Hauff. Exploring Collaborative Navigation Support in Collaborative Product Search. In the 2021 SIGIR Workshop On eCommerce.

2. Sheikh Muhammad Sarwar, Felipe Moraes, Jiepu Jiang, and James Allan. Utility of Missing Concepts in Query-biased Summarization. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2056-2060. 2021.

3. Felipe Moraes, Jie Yang, Rongting Zhang and Vanessa Murdock. The Role of Attributes in Product Quality Comparisons. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pp. 253-262. 2020.

4. Nirmal Roy, Felipe Moraes and Claudia Hauff. Exploring Users' Learning Gains within Search Sessions. In Proceedings of the 2020 conference on human information interaction and retrieval, pp. 432-436. 2020.

5. Felipe Moraes, Kilian Grashoff and Claudia Hauff. On the Impact of Group Size on Collaborative Search Effectiveness. Information Retrieval Journal 22, no. 5 (2019): 476-498.

6. Felipe Moraes and Claudia Hauff. node-indri: moving the Indri toolkit to the modern Web stack. In European Conference on Information Retrieval, pp. 241-245. Springer, Cham, 2019.

7. Daan Rennings, Felipe Moraes and Claudia Hauff. An Axiomatic Approach to Diagnosing Neural IR Models. In European Conference on Information Retrieval, pp. 489-503. Springer, Cham, 2019.

8. Felipe Moraes, Sindunuraga Rikarno Putra and Claudia Hauff. Contrasting search as a learning activity with instructor-designed learning. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 167-176. 2018.

9. Sindunuraga Rikarno Putra, Kilian Grashoff, Felipe Moraes and Claudia Hauff. On the Development of a Collaborative Search System. In Proceedings of the 1st Design of Experimental Search & Information REtrieval Systems, pp. 76-82. 2018.

10. Sindunuraga Rikarno Putra*, Felipe Moraes* and Claudia Hauff. SearchX: Empowering Collaborative Search Research. In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 1265-1268. 2018. (* Joint first authors)

11. Felipe Moraes, Rodrygo L.T. Santos, and Nivio Ziviani. On Effective Dynamic Search in Specialized Domains. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, pp. 177-184. 2017.

12. Felipe Moraes, Mário S. Alvim, and Rodrygo L.T. Santos. Modeling Information Flow in Dynamic Information Retrieval. In Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval, pp. 305-308. 2017.

13. Felipe Moraes, Rodrygo L.T. Santos, and Nivio Ziviani. UFMG at the TREC 2016 Dynamic Domain track. In Proceedings of the 25th Text REtrieval Conference, 2016.

14. Felipe Moraes, Marisa Vasconcelos, Patrick Prado, Daniel Dalip, Jussara Almeida, and Marcos Gonçalves. Polarity Detection of Foursquare Tips. In International Conference on Social Informatics, pp. 153-162. Springer, Cham, 2013.

15. Felipe Moraes, Marisa Vasconcelos, Patrick Prado, Jussara Almeida, and Marcos Gonçalves. Polarity Analysis of Micro Reviews in Foursquare. In Proceedings of the 19th Brazilian symposium on multimedia and the web, pp. 113-120. 2013.

SIKS Dissertation Series

Since 1998, all dissertations written by Ph.D. students who have conducted their research under auspices of a senior research fellow of the SIKS research school are published in the SIKS Dissertation Series.

——1998——

1998-01 Johan van den Akker, *DEGAS - An Active, Temporal Database of Autonomous Objects*, CWI

1998-02 Floris Wiesman, *Information Retrieval by Graphically Browsing Meta-Information*, UM

1998-03 Ans Steuten, *A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective*, TUD

1998-04 Dennis Breuker, *Memory versus Search in Games*, UM

1998-05 E.W.Oskamp, *Computerondersteuning bij Straftoemeting*, RUL

——1999——

1999-01 Mark Sloof, *Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products*, VU

1999-02 Rob Potharst, *Classification Using Decision Trees and Neural Nets*, EUR

1999-03 Don Beal, *The Nature of Minimax Search*, UM

1999-04 Jacques Penders, *The practical Art of Moving Physical Objects*, UM

1999-05 Aldo de Moor, *Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems*, KUB

1999-06 Niek J.E. Wijngaards, *Re-design of Compositional Systems*, VU

1999-07 David Spelt, *Verification Support for Object Database Design*, UT

1999-08 Jacques H.J. Lenting, *Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation*, UM

——2000——

2000-01 Frank Niessink, *Perspectives on Improving Software Maintenance*, VU

2000-02 Koen Holtman, *Prototyping of CMS Storage Management*, TUE

2000-03 Carolien M.T. Metselaar, *Sociaal-organisatorische Gevolgen van Kennistechnologie; een Procesbenadering en Actorperspectief*, UvA

2000-04 Geert de Haan, *ETAG, A Formal Model of Competence Knowledge for User Interface Design*, VU

2000-05 Ruud van der Pol, *Knowledge-Based Query Formulation in Information Retrieval*, UM

2000-06 Rogier van Eijk, *Programming Languages for Agent Communication*, UU

2000-07 Niels Peek, *Decision-Theoretic Planning of Clinical Patient Management*, UU

2000-08 Veerle Coup, *Sensitivity Analyis of Decision-Theoretic Networks*, EUR

2000-09 Florian Waas, *Principles of Probabilistic Query Optimization*, CWI

2000-10 Niels Nes, *Image Database Management System Design Considerations, Algorithms, and Architecture*, CWI

2000-11 Jonas Karlsson, *Scalable Distributed Data Structures for Database Management*, CWI

——2001——

2001-01 Silja Renooij, *Qualitative Approaches to Quantifying Probabilistic Networks*, UU

2001-02 Koen Hindriks, *Agent Programming Languages: Programming with Mental Models*, UU

2001-03 Maarten van Someren, *Learning as Problem Solving*, UvA

2001-04 Evgueni Smirnov, *Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets*, UM

2001-05 Jacco van Ossenbruggen, *Processing Structured Hypermedia: A Matter of Style*, VU

2001-06 Martijn van Welie, *Task-based User Interface Design*, VU

2001-07 Bastiaan Schonhage, *Diva: Architectural Perspectives on Information Visualization*, VU

2001-08 Pascal van Eck, *A Compositional Semantic Structure for Multi-Agent Systems Dynamics*, VU

2001-09 Pieter Jan 't Hoen, *Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes*, RUL

2001-10 Maarten Sierhuis, *Modeling and Simulating Work Practice BRAHMS: a Multi-Agent Modeling and Simulation Language for Work Practice Analysis and Design*, UvA

2001-11 Tom M. van Engers, *Knowledge Management: The Role of Mental Models in Business Systems Design*, VU

——2002——

2002-01 Nico Lassing, *Architecture-Level Modifiability Analysis*, VU

2002-02 Roelof van Zwol, *Modelling and Searching Web-Based Document Collections*, UT

2002-03 Henk Ernst Blok, *Database Optimization Aspects for Information Retrieval*, UT

2002-04 Juan Roberto Castelo Valdueza, *The Discrete Acyclic Digraph Markov Model in Data Mining*, UU

2002-05 Radu Serban, *The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents*, VU

2002-06 Laurens Mommers, *Applied Legal Epistemology; Building a Knowledge-Based Ontology of the Legal Domain*, UL

2002-07 Peter Boncz, *Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications*, CWI

2002-08 Jaap Gordijn *Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas*, CWI

2002-09 Willem-Jan van den Heuvel, *Integrating Modern Business Applications with Objectified Legacy Systems*, KUB

2002-10 Brian Sheppard *Towards Perfect Play of Scrabble*, UM

2002-11 Wouter C.A. Wijngaards, *Agent Based Modelling of Dynamics: Biological and Organisational Applications*, VU

2002-12 Albrecht Schmidt, *Processing XML in Database Systems*, UvA

2002-13 Hongjing Wu, *A Reference Architecture for Adaptive Hypermedia Applications*, TUE

2002-14 Wieke de Vries, *Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems*, UU

2002-15 Rik Eshuis, *Semantics and Verification of UML Activity Diagrams for Workflow Modelling*, UT

2002-16 Pieter van Langen, *The Anatomy of Design: Foundations, Models and Applications*, VU

2002-17 Stefan Manegold, *Understanding, Modeling, and Improving Main-Memory Database Performance*, UvA

——2003——

2003-01 Heiner Stuckenschmidt, *Ontology-Based Information Sharing in Weakly Structured Environments*, VU

2003-02 Jan Broersen, *Modal Action Logics for Reasoning About Reactive Systems*, VU

2003-03 Martijn Schuemie, *Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy*, TUD

2003-04 Milan Petkovic, *Content-Based Video Retrieval Supported by Database Technology*, UT

2003-05 Jos Lehmann, *Causation in Artificial Intelligence and Law - A modelling approach*, UvA

2003-06 Boris van Schooten, *Development and Specification of Virtual Environments*, UT

2003-07 Machiel Jansen, *Formal Explorations of Knowledge Intensive Tasks*, UvA

2003-08 Yongping Ran, *Repair Based Scheduling*, UM

2003-09 Rens Kortmann, *The Resolution of Visually Guided Behaviour*, UM

2003-10 Andreas Lincke, *Electronic Business Negotiation: Some Experimental Studies on the Interaction Between Medium, Innovation Context and Culture*, UvT

2003-11 Simon Keizer, *Reasoning Under Uncertainty in Natural Language Dialogue Using Bayesian Networks*, UT

2003-12 Roeland Ordelman, *Dutch Speech Recognition in Multi-Media Information Retrieval*, UT

2003-13 Jeroen Donkers, *Nosce Hostem - Searching with Opponent Models*, UM

2003-14 Stijn Hoppenbrouwers, *Freezing Language: Conceptualisation Processes across ICT-Supported Organisations*, KUN

2003-15 Mathijs de Weerdt, *Plan Merging in Multi-Agent Systems*, TUD

2003-16 Menzo Windhouwer, *Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses*, CWI

2003-17 David Jansen, *Extensions of Statecharts with Probability, Time, and Stochastic Timing*, UT

2003-18 Levente Kocsis, *Learning Search Decisions*, UM

——2004——

2004-01 Virginia Dignum, *A Model for Organizational Interaction: Based on Agents, Founded in Logic*, UU

2004-02 Lai Xu, *Monitoring Multi-party Contracts for E-business*, UvT

2004-03 Perry Groot, *A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving*, VU

2004-04 Chris van Aart, *Organizational Principles for Multi-Agent Architectures*, UvA

2004-05 Viara Popova, *Knowledge Discovery and Monotonicity*, EUR

2004-06 Bart-Jan Hommes, *The Evaluation of Business Process Modeling Techniques*, TUD

2004-07 Elise Boltjes, *Voorbeeldig Onderwijs; Voorbeeldgestuurd Onderwijs, een Opstap naar Abstract Denken, Vooral voor Meisjes*, UM

2004-08 Joop Verbeek, *Politie en de Nieuwe Internationale Informatiemarkt: Grensregionale Politiële Gegevensuitwisseling en Digitale Expertise*, UM

2004-09 Martin Caminada, *For the Sake of the Argument; Explorations into Argument-Based Reasoning*, VU

2004-10 Suzanne Kabel, *Knowledge-rich Indexing of Learning-Objects*, UvA

2004-11 Michel Klein, *Change Management for Distributed Ontologies*, VU

2004-12 The Duy Bui, *Creating Emotions and Facial Expressions for Embodied Agents*, UT

2004-13 Wojciech Jamroga, *Using Multiple Models of Reality: On Agents who Know how to Play*, UT

2004-14 Paul Harrenstein, *Logic in Conflict. Logical Explorations in Strategic Equilibrium*, UU

2004-15 Arno Knobbe, *Multi-Relational Data Mining*, UU

2004-16 Federico Divina, *Hybrid Genetic Relational Search for Inductive Learning*, UvA

2004-17 Mark Winands, *Informed Search in Complex Games*, UM

2004-18 Vania Bessa Machado, *Supporting the Construction of Qualitative Knowledge Models*, UvA

2004-19 Thijs Westerveld, *Using Generative Probabilistic Models for Multi-Media Retrieval*, UT

2004-20 Madelon Evers, *Learning from Design: Facilitating Multidisciplinary Design Teams*, Nyenrode

——2005——

2005-01 Floor Verdenius, *Methodological Aspects of Designing Induction-Based Applications*, UvA

2005-02 Erik van der Werf, *AI Techniques for the Game of Go*, UM

2005-03 Franc Grootjen, *A Pragmatic Approach to the Conceptualisation of Language*, RUN

2005-04 Nirvana Meratnia, *Towards Database Support for Moving Object data*, UT

2005-05 Gabriel Infante-Lopez, *Two-Level Probabilistic Grammar for Natural Language Parsing*, UvA

2005-06 Pieter Spronck, *Adaptive Game AI*, UM

2005-07 Flavius Frasincar, *Hypermedia Presentation Generation for Semantic Web Information Systems*, TUE

2005-08 Richard Vdovjak, *A Model-Driven Approach for Building Distributed Ontology-Based Web Applications*, TUE

2005-09 Jeen Broekstra, *Storage, Querying and Inferencing for Semantic Web Languages*, VU

2005-10 Anders Bouwer, *Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments*, UvA

2005-11 Elth Ogston, *Agent Based Matchmaking and Clustering - A Decentralized Approach to Search*, VU

2005-12 Csaba Boer, *Distributed Simulation in Industry*, EUR

2005-13 Fred Hamburg, *Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen*, UL

2005-14 Borys Omelayenko, *Web-Service Configuration on the Semantic Web; Exploring how Semantics Meets Pragmatics*, VU

2005-15 Tibor Bosse, *Analysis of the Dynamics of Cognitive Processes*, VU

2005-16 Joris Graaumans, *Usability of XML Query Languages*, UU

2005-17 Boris Shishkov, *Software Specification Based on Reusable Business Components*, TUD

2005-18 Danielle Sent, *Test-Selection Strategies for Probabilistic Networks*, UU

2005-19 Michel van Dartel, *Situated Representation*, UM

2005-20 Cristina Coteanu, *Cyber Consumer Law, State of the Art and Perspectives*, UL

2005-21 Wijnand Derks, *Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics*, UT

——2006——

2006-01 Samuil Angelov, *Foundations of B2B Electronic Contracting*, TUE

2006-02 Cristina Chisalita, *Contextual Issues in the Design and Use of Information Technology in Organizations*, VU

2006-03 Noor Christoph, *The Role of Metacognitive Skills in Learning to Solve Problems*, UvA

2006-04 Marta Sabou, *Building Web Service Ontologies*, VU

2006-05 Cees Pierik, *Validation Techniques for Object-Oriented Proof Outlines*, UU

2006-06 Ziv Baida, *Software-Aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling*, VU

2006-07 Marko Smiljanic, *XML Schema Matching - Balancing Efficiency and Effectiveness by Means of Clustering*, UT

2006-08 Eelco Herder, *Forward, Back and Home Again - Analyzing User Behavior on the Web*, UT

2006-09 Mohamed Wahdan, *Automatic Formulation of the Auditor's Opinion*, UM

2006-10 Ronny Siebes, *Semantic Routing in Peer-to-Peer Systems*, VU

2006-11 Joeri van Ruth, *Flattening Queries over Nested Data Types*, UT

2006-12 Bert Bongers, *Interactivation - Towards an E-cology of People, our Technological Environment, and the Arts*, VU

2006-13 Henk-Jan Lebbink, *Dialogue and Decision Games for Information Exchanging Agents*, UU

2006-14 Johan Hoorn, *Software Requirements: Update, Upgrade, Re-design - Towards a Theory of Requirements Change*, VU

2006-15 Rainer Malik, *CONAN: Text Mining in the Biomedical Domain*, UU

2006-16 Carsten Riggelsen, *Approximation Methods for Efficient Learning of Bayesian Networks*, UU

2006-17 Stacey Nagata, *User Assistance for Multitasking with Interruptions on a Mobile Device*, UU

2006-18 Valentin Zhizhkun, *Graph Transformation for Natural Language Processing*, UvA

2006-19 Birna van Riemsdijk, *Cognitive Agent Programming: A Semantic Approach*, UU

2006-20 Marina Velikova, *Monotone Models for Prediction in Data Mining*, UvT

2006-21 Bas van Gils, *Aptness on the Web*, RUN

2006-22 Paul de Vrieze, *Fundaments of Adaptive Personalisation*, RUN

2006-23 Ion Juvina, *Development of Cognitive Model for Navigating on the Web*, UU

2006-24 Laura Hollink, *Semantic Annotation for Retrieval of Visual Resources*, VU

2006-25 Madalina Drugan, *Conditional Log-Likelihood MDL and Evolutionary MCMC*, UU

2006-26 Vojkan Mihajlovic, *Score Region Algebra: A Flexible Framework for Structured Information Retrieval*, UT

2006-27 Stefano Bocconi, *Vox Populi: Generating Video Documentaries from Semantically Annotated Media Repositories*, CWI

2006-28 Borkur Sigurbjornsson, *Focused Information Access using XML Element Retrieval*, UvA

——2007——

2007-01 Kees Leune, *Access Control and Service-Oriented Architectures*, UvT

2007-02 Wouter Teepe, *Reconciling Information Exchange and Confidentiality: A Formal Approach*, RUG

2007-03 Peter Mika, *Social Networks and the Semantic Web*, VU

2007-04 Jurriaan van Diggelen, *Achieving Semantic Interoperability in Multi-Agent Systems: a Dialogue-Based Approach*, UU

2007-05 Bart Schermer, *Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-Enabled Surveillance*, UL

2007-06 Gilad Mishne, *Applied Text Analytics for Blogs*, UvA

2007-07 Natasa Jovanovic, *To Whom It May Concern - Addressee Identification in Face-to-Face Meetings*, UT

2007-08 Mark Hoogendoorn, *Modeling of Change in Multi-Agent Organizations*, VU

2007-09 David Mobach, *Agent-Based Mediated Service Negotiation*, VU

2007-10 Huib Aldewereld, *Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols*, UU

2007-11 Natalia Stash, *Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System*, TUE

2007-12 Marcel van Gerven, *Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty*, RUN

2007-13 Rutger Rienks, *Meetings in Smart Environments; Implications of Progressing Technology*, UT

2007-14 Niek Bergboer, *Context-Based Image Analysis*, UM

2007-15 Joyca Lacroix, *NIM: a Situated Computational Memory Model*, UM

2007-16 Davide Grossi, *Designing Invisible Handcuffs. Formal Investigations in Institutions and Organizations for Multi-Agent Systems*, UU

2007-17 Theodore Charitos, *Reasoning with Dynamic Networks in Practice*, UU

2007-18 Bart Orriens, *On the Development an Management of Adaptive Business Collaborations*, UvT

2007-19 David Levy, *Intimate Relationships with Artificial Partners*, UM

2007-20 Slinger Jansen, *Customer Configuration Updating in a Software Supply Network*, UU

2007-21 Karianne Vermaas, *Fast Diffusion and Broadening Use: A Research on Residential Adoption and Usage of Broadband Internet in the Netherlands Between 2001 and 2005*, UU

2007-22 Zlatko Zlatev, *Goal-Oriented Design of Value and Process Models from Patterns*, UT

2007-23 Peter Barna, *Specification of Application Logic in Web Information Systems*, TUE

2007-24 Georgina Ram�rez Camps, *Structural Features in XML Retrieval*, CWI

2007-25 Joost Schalken, *Empirical Investigations in Software Process Improvement*, VU

——2008——

2008-01 Katalin Boer-Sorb�n, *Agent-Based Simulation of Financial Markets: A Modular, Continuous-Time Approach*, EUR

2008-02 Alexei Sharpanskykh, *On Computer-Aided Methods for Modeling and Analysis of Organizations*, VU

2008-03 Vera Hollink, *Optimizing Hierarchical Menus: a Usage-Based Approach*, UvA

2008-04 Ander de Keijzer, *Management of Uncertain Data - Towards Unattended Integration*, UT

2008-05 Bela Mutschler, *Modeling and Simulating Causal Dependencies on Process-Aware Information Systems from a Cost Perspective*, UT

2008-06 Arjen Hommersom, *On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective*, RUN

2008-07 Peter van Rosmalen, *Supporting the Tutor in the Design and Support of Adaptive E-Learning*, OU

2008-08 Janneke Bolt, *Bayesian Networks: Aspects of Approximate Inference*, UU

2008-09 Christof van Nimwegen, *The Paradox of the Guided User: Assistance can be Counter-Effective*, UU

2008-10 Wauter Bosma, *Discourse Oriented Summarization*, UT

2008-11 Vera Kartseva, *Designing Controls for Network Organizations: A Value-Based Approach*, VU

2008-12 Jozsef Farkas, *A Semiotically Oriented Cognitive Model of Knowledge Representation*, RUN

2008-13 Caterina Carraciolo, *Topic Driven Access to Scientific Handbooks*, UvA

2008-14 Arthur van Bunningen, *Context-Aware Querying; Better Answers with Less Effort*, UT

2008-15 Martijn van Otterlo, *The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains*, UT

2008-16 Henriette van Vugt, *Embodied Agents from a User's Perspective*, VU

2008-17 Martin Op 't Land, *Applying Architecture and Ontology to the Splitting and Allying of Enterprises*, TUD

2008-18 Guido de Croon, *Adaptive Active Vision*, UM

2008-19 Henning Rode, *From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search*, UT

2008-20 Rex Arendsen, *Geen Bericht, Goed Bericht. Een Onderzoek naar de Effecten van de Introductie van Elektronisch Berichtenverkeer met de Overheid op de Administratieve Lasten van Bedrijven*, UvA

2008-21 Krisztian Balog, *People Search in the Enterprise*, UvA

2008-22 Henk Koning, *Communication of IT-Architecture*, UU

2008-23 Stefan Visscher, *Bayesian Network Models for the Management of Ventilator-Associated Pneumonia*, UU

2008-24 Zharko Aleksovski, *Using Background Knowledge in Ontology Matching*, VU

2008-25 Geert Jonker, *Efficient and Equitable Exchange in Air Traffic Management Plan Repair Using Spender-Signed Currency*, UU

2008-26 Marijn Huijbregts, *Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled*, UT

2008-27 Hubert Vogten, *Design and Implementation Strategies for IMS Learning Design*, OU

2008-28 Ildiko Flesch, *On the Use of Independence Relations in Bayesian Networks*, RUN

2008-29 Dennis Reidsma, *Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans*, UT

2008-30 Wouter van Atteveldt, *Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content*, VU

2008-31 Loes Braun, *Pro-Active Medical Information Retrieval*, UM

2008-32 Trung H. Bui, *Toward Affective Dialogue Management using Partially Observable Markov Decision Processes*, UT

2008-33 Frank Terpstra, *Scientific Workflow Design; Theoretical and Practical Issues*, UvA

2008-34 Jeroen de Knijf, *Studies in Frequent Tree Mining*, UU

2008-35 Ben Torben Nielsen, *Dendritic Morphologies: Function Shapes Structure*, UvT

—2009—

2009-01 Rasa Jurgelenaite, *Symmetric Causal Independence Models*, RUN

2009-02 Willem Robert van Hage, *Evaluating Ontology-Alignment Techniques*, VU

2009-03 Hans Stol, *A Framework for Evidence-Based Policy Making Using IT*, UvT

2009-04 Josephine Nabukenya, *Improving the Quality of Organisational Policy Making Using Collaboration Engineering*, RUN

2009-05 Sietse Overbeek, *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*, RUN

2009-06 Muhammad Subianto, *Understanding Classification*, UU

2009-07 Ronald Poppe, *Discriminative Vision-Based Recovery and Recognition of Human Motion*, UT

2009-08 Volker Nannen, *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*, VU

2009-09 Benjamin Kanagwa, *Design, Discovery and Construction of Service-Oriented Systems*, RUN

2009-10 Jan Wielemaker, *Logic Programming for Knowledge-Intensive Interactive Applications*, UvA

2009-11 Alexander Boer, *Legal Theory, Sources of Law & the Semantic Web*, UvA

2009-12 Peter Massuthe, *Operating Guidelines for Services*, TUE/Humboldt-Universitaet zu Berlin

2009-13 Steven de Jong, *Fairness in Multi-Agent Systems*, UM

2009-14 Maksym Korotkiy, *From Ontology-Enabled Services to Service-Enabled Ontologies (Making Ontologies Work in E-Science with ONTO-SOA)*, VU

2009-15 Rinke Hoekstra, *Ontology Representation - Design Patterns and Ontologies that Make Sense*, UvA

2009-16 Fritz Reul, *New Architectures in Computer Chess*, UvT

2009-17 Laurens van der Maaten, *Feature Extraction from Visual Data*, UvT

2009-18 Fabian Groffen, *Armada, An Evolving Database System*, CWI

2009-19 Valentin Robu, *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*, CWI

2009-20 Bob van der Vecht, *Adjustable Autonomy: Controling Influences on Decision Making*, UU

2009-21 Stijn Vanderlooy, *Ranking and Reliable Classification*, UM

2009-22 Pavel Serdyukov, *Search For Expertise: Going Beyond Direct Evidence*, UT

2009-23 Peter Hofgesang, *Modelling Web Usage in a Changing Environment*, VU

2009-24 Annerieke Heuvelink, *Cognitive Models for Training Simulations*, VU

2009-25 Alex van Ballegooij, *RAM: Array Database Management Through Relational Mapping*, CWI

2009-26 Fernando Koch, *An Agent-Based Model for the Development of Intelligent Mobile Services*, UU

2009-27 Christian Glahn, *Contextual Support of Social Engagement and Reflection on the Web*, OU

2009-28 Sander Evers, *Sensor Data Management with Probabilistic Models*, UT

2009-29 Stanislav Pokraev, *Model-Driven Semantic Integration of Service-Oriented Applications*, UT

2009-30 Marcin Zukowski, *Balancing Vectorized Query Execution with Bandwidth-Optimized Storage*, CWI

2009-31 Sofiya Katrenko, *A Closer Look at Learning Relations from Text*, UvA

2009-32 Rik Farenhorst, and Remco de Boer, *Architectural Knowledge Management: Supporting Architects and Auditors*, VU

2009-33 Khiet Truong, *How Does Real Affect Affect Affect Recognition In Speech?*, UT

2009-34 Inge van de Weerd, *Advancing in Software Product Management: An Incremental Method Engineering Approach*, UU

2009-35 Wouter Koelewijn, *Privacy en Politiegegevens; Over Geautomatiseerde Normatieve Informatie-Uitwisseling*, UL

2009-36 Marco Kalz, *Placement Support for Learners in Learning Networks*, OU

2009-37 Hendrik Drachsler, *Navigation Support for Learners in Informal Learning Networks*, OU

2009-38 Riina Vuorikari, *Tags and Self-Organisation: a Metadata E-cology for Learning Resources in a Multilingual Context*, OU

2009-39 Christian Stahl, *Service Substitution - A Behavioral Approach Based on Petri Nets*, TUE/Humboldt-Universitaet zu Berlin

2009-40 Stephan Raaijmakers, *Multinomial Language Learning: Investigations into the Geometry of Language*, UvT

2009-41 Igor Berezhnyy, *Digital Analysis of Paintings*, UvT

2009-42 Toine Bogers, *Recommender Systems for Social Bookmarking*, UvT

2009-43 Virginia Nunes Leal Franqueira, *Finding Multi-Step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*, UT

2009-44 Roberto Santana Tapia, *Assessing Business-IT Alignment in Networked Organizations*, UT

2009-45 Jilles Vreeken, *Making Pattern Mining Useful*, UU

2009-46 Loredana Afanasiev, *Querying XML: Benchmarks and Recursion*, UvA

—2010—

2010-01 Matthijs van Leeuwen, *Patterns that Matter*, UU

2010-02 Ingo Wassink, *Work flows in Life Science*, UT

2010-03 Joost Geurts, *A Document Engineering Model and Processing Framework for Multimedia documents*, CWI

2010-04 Olga Kulyk, *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*, UT

2010-05 Claudia Hauff, *Predicting the Effectiveness of Queries and Retrieval Systems*, UT

2010-06 Sander Bakkes, *Rapid Adaptation of Video Game AI*, UvT

2010-07 Wim Fikkert, *Gesture Interaction at a Distance*, UT

2010-08 Krzysztof Siewicz, *Towards an Improved Regulatory Framework of Free Software. Protecting User Freedoms in a World of Software Communities and E-Governments*, UL

2010-09 Hugo Kielman, *Politiële Gegevensverwerking en Privacy, Naar een Effectieve Waarborging*, UL

2010-10 Rebecca Ong, *Mobile Communication and Protection of Children*, UL

2010-11 Adriaan Ter Mors, *The World According to MARP: Multi-Agent Route Planning*, TUD

2010-12 Susan van den Braak, *Sensemaking Software for Crime Analysis*, UU

2010-13 Gianluigi Folino, *High Performance Data Mining Using Bio-inspired techniques*, RUN

2010-14 Sander van Splunter, *Automated Web Service Reconfiguration*, VU

2010-15 Lianne Bodenstaff, *Managing Dependency Relations in Inter-Organizational Models*, UT

2010-16 Sicco Verwer, *Efficient Identification of Timed Automata, Theory and Practice*, TUD

2010-17 Spyros Kotoulas, *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*, VU

2010-18 Charlotte Gerritsen, *Caught in the Act: Investigating Crime by Agent-Based Simulation*, VU

2010-19 Henriette Cramer, *People's Responses to Autonomous and Adaptive Systems*, UvA

2010-20 Ivo Swartjes, *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*, UT

2010-21 Harold van Heerde, *Privacy-Aware Data Management by Means of Data Degradation*, UT

2010-22 Michiel Hildebrand, *End-user Support for Access to Heterogeneous Linked Data*, CWI

2010-23 Bas Steunebrink, *The Logical Structure of Emotions*, UU

2010-24 Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies*

2010-25 Zulfiqar Ali Memon, *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*, VU

2010-26 Ying Zhang, *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*, CWI

2010-27 Marten Voulon, *Automatisch Contracteren*, UL

2010-28 Arne Koopman, *Characteristic Relational Patterns*, UU

2010-29 Stratos Idreos, *Database Cracking: Towards Auto-Tuning Database Kernels*, CWI

2010-30 Marieke van Erp, *Accessing Natural History - Discoveries in Data Cleaning, Structuring, and Retrieval*, UvT

2010-31 Victor de Boer, *Ontology Enrichment from Heterogeneous Sources on the Web*, UvA

2010-32 Marcel Hiel, *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*, UvT

2010-33 Robin Aly, *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*, UT

2010-34 Teduh Dirgahayu, *Interaction Design in Service Compositions*, UT

2010-35 Dolf Trieschnigg, *Proof of Concept: Concept-based Biomedical Information Retrieval*, UT

2010-36 Jose Janssen, *Paving the Way for Lifelong Learning; Facilitating Competence Development Through a Learning Path Specification*, OU

2010-37 Niels Lohmann, *Correctness of Services and Their Composition*, TUE

2010-38 Dirk Fahland, *From Scenarios to Components*, TUE

2010-39 Ghazanfar Farooq Siddiqui, *Integrative Modeling of Emotions in Virtual Agents*, VU

2010-40 Mark van Assem, *Converting and Integrating Vocabularies for the Semantic Web*, VU

2010-41 Guillaume Chaslot, *Monte-Carlo Tree Search*, UM

2010-42 Sybren de Kinderen, *Needs-Driven Service Bundling in a Multi-Supplier Setting - the Computational E3-Service Approach*, VU

2010-43 Peter van Kranenburg, *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*, UU

2010-44 Pieter Bellekens, *An Approach Towards Context-Sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*, TUE

2010-45 Vasilios Andrikopoulos, *A Theory and Model for the Evolution of Software Services*, UvT

2010-46 Vincent Pijpers, *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*, VU

2010-47 Chen Li, *Mining Process Model Variants: Challenges, Techniques, Examples*, UT

2010-48 Milan Lovric, *Behavioral Finance and Agent-Based Artificial Markets*, EUR

2010-49 Jahn-Takeshi Saito, *Solving Difficult Game Positions*, UM

2010-50 Bouke Huurnink, *Search in Audiovisual Broadcast Archives*, UvA

2010-51 Alia Khairia Amin, *Understanding and Supporting Information Seeking Tasks in Multiple Sources*, CWI

2010-52 Peter-Paul van Maanen , *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*, VU

2010-53 Edgar Meij, *Combining Concepts and Language Models for Information Access*, UvA

—–2011—–

2011-01 Botond Cseke, *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*, RUN

2011-02 Nick Tinnemeier, *Work Flows in Life Science*, UU

2011-03 Jan Martijn van der Werf, *Compositional Design and Verification of Component-Based Information Systems*, TUE

2011-04 Hado van Hasselt, *Insights in Reinforcement Learning; Formal Analysis and Empirical Evaluation of Temporal-Difference Learning Algorithms*, UU

2011-05 Base van der Raadt, *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.*, VU

2011-06 Yiwen Wang, *Semantically-Enhanced Recommendations in Cultural Heritage*, TUE

2011-07 Yujia Cao, *Multimodal Information Presentation for High Load Human Computer Interaction*, UT

2011-08 Nieske Vergunst, *BDI-Based Generation of Robust Task-Oriented Dialogues*, UU

2011-09 Tim de Jong, *Contextualised Mobile Media for Learning*, OU

2011-10 Bart Bogaert, *Cloud Content Contention*, UvT

2011-11 Dhaval Vyas, *Designing for Awareness: An Experience-Focused HCI Perspective*, UT

2011-12 Carmen Bratosin, *Grid Architecture for Distributed Process Mining*, TUE

2011-13 Xiaoyu Mao, *Airport Under Control. Multi-Agent Scheduling for Airport Ground Handling*, UvT

2011-14 Milan Lovric, *Behavioral Finance and Agent-Based Artificial Markets*, EUR

2011-15 Marijn Koolen, *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*, UvA

2011-16 Maarten Schadd, *Selective Search in Games of Different Complexity*, UM

2011-17 Jiyin He, *Exploring Topic Structure: Coherence, Diversity and Relatedness*, UvA

2011-18 Mark Ponsen, *Strategic Decision-Making in Complex Games*, UM

2011-19 Ellen Rusman, *The Mind ' s Eye on Personal Profiles*, OU

2011-20 Qing Gu, *Guiding Service-Oriented Software Engineering - A View-Based Approach*, VU

2011-21 Linda Terlouw, *Modularization and Specification of Service-Oriented Systems*, TUD

2011-22 Junte Zhang, *System Evaluation of Archival Description and Access*, UvA

2011-23 Wouter Weerkamp, *Finding People and their Utterances in Social Media*, UvA

2011-24 Herwin van Welbergen, *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*, UT

2011-25 Syed Waqar ul Qounain Jaffry, *Analysis and Validation of Models for Trust Dynamics*, VU

2011-26 Matthijs Aart Pontier, *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*, VU

2011-27 Aniel Bhulai, *Dynamic Website Optimization Through Autonomous Management of Design Patterns*, VU

2011-28 Rianne Kaptein, *Effective Focused Retrieval by Exploiting Query Context and Document Structure*, UvA

2011-29 Faisal Kamiran, *Discrimination-Aware Classification*, TUE

2011-30 Egon van den Broek, *Affective Signal Processing (ASP): Unraveling the Mystery of Emotions*, UT

2011-31 Ludo Waltman, *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, EUR

2011-32 Nees-Jan van Eck, *Methodological Advances in Bibliometric Mapping of Science*, EUR

2011-33 Tom van der Weide, *Arguing to Motivate Decisions*, UU

2011-34 Paolo Turrini, *Strategic Reasoning in Interdependence: Logical and Game-Theoretical Investigations*, UU

2011-35 Maaike Harbers, *Explaining Agent Behavior in Virtual Training*, UU

2011-36 Erik van der Spek, *Experiments in Serious Game Design: a Cognitive Approach*, UU

2011-37 Adriana Burlutiu, *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference*, RUN

2011-38 Nyree Lemmens, *Bee-inspired Distributed Optimization*, UM

2011-39 Joost Westra, *Organizing Adaptation Using Agents in Serious Games*, UU

2011-40 Viktor Clerc, *Architectural Knowledge Management in Global Software Development*, VU

2011-41 Luan Ibraimi, *Cryptographically Enforced Distributed Data Access Control*, UT

2011-42 Michal Sindlar, *Explaining Behavior Through Mental State Attribution*, UU

2011-43 Henk van der Schuur, *Process Improvement Through Software Operation Knowledge*, UU

2011-44 Boris Reuderink, *Robust Brain-Computer Interfaces*, UT

2011-45 Herman Stehouwer, *Statistical Language Models for Alternative Sequence Selection*, UvT

2011-46 Beibei Hu, *Towards Contextualized Information Delivery: A Rule-Based Architecture for the Domain of Mobile Police Work*, TUD

2011-47 Azizi Bin Ab Aziz, *Exploring Computational Models for Intelligent Support of Persons with Depression*, VU

2011-48 Mark Ter Maat, *Response Selection and Turn-Taking for a Sensitive Artificial Listening Agent*, UT

2011-49 Andreea Niculescu, *Conversational Interfaces for Task-Oriented Spoken Dialogues: Design Aspects Influencing Interaction Quality*, UT

——2012——

2012-01 Terry Kakeeto, *Relationship Marketing for SMEs in Uganda*, UvT

2012-02 Muhammad Umair, *Adaptivity, Emotion, and Rationality in Human and Ambient Agent Models*, VU

2012-03 Adam Vanya, *Supporting Architecture Evolution by Mining Software Repositories*, VU

2012-04 Jurriaan Souer, *Development of Content Management System-based Web Applications*, UU

2012-05 Marijn Plomp, *Maturing Inter-Organisational Information Systems*, UU

2012-06 Wolfgang Reinhardt, *Awareness Support for Knowledge Workers in Research Networks*, OU

2012-07 Rianne van Lambalgen, *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*, VU

2012-08 Gerben de Vries, *Kernel Methods for Vessel Trajectories*, UvA

2012-09 Ricardo Neisse, *Trust and Privacy Management Support for Context-Aware Service Platforms*, UT

2012-10 David Smits, *Towards a Generic Distributed Adaptive Hypermedia Environment*, TUE

2012-11 J.C.B. Rantham Prabhakara, *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*, TUE

2012-12 Kees van der Sluijs, *Model Driven Design and Data Integration in Semantic Web Information Systems*, TUE

2012-13 Suleman Shahid, *Fun and Face: Exploring Non-Verbal Expressions of Emotion During Playful Interactions*, UvT

2012-14 Evgeny Knutov(TUE) *Generic Adaptation Framework for Unifying Adaptive Web-based Systems*, TUE

2012-15 Natalie van der Wal, *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes*, VU

2012-16 Fiemke Both, *Helping People by Understanding Them - Ambient Agents Supporting Task Execution and Depression Treatment*, VU

2012-17 Amal Elgammal, *Towards a Comprehensive Framework for Business Process Compliance*, UvT

2012-18 Eltjo Poort, *Improving Solution Architecting Practices*, VU

2012-19 Helen Schonenberg, *What's Next? Operational Support for Business Process Execution*, TUE

2012-20 Ali Bahramisharif, *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing*, RUN

2012-21 Roberto Cornacchia, *Querying Sparse Matrices for Information Retrieval*, TUD

2012-22 Thijs Vis, *Intelligence, Politie en Veiligheidsdienst: Verenigbare Grootheden?*, UvT

2012-23 Christian Muehl, *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction*, UT

2012-24 Laurens van der Werff, *Evaluation of Noisy Transcripts for Spoken Document Retrieval*, UT

2012-25 Silja Eckartz, *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application*, UT

2012-26 Emile de Maat, *Making Sense of Legal Text*, UvA

2012-27 Hayrettin Gurkok, *Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games*, UT

2012-28 Nancy Pascall, *Engendering Technology Empowering Women*, UvT

2012-29 Almer Tigelaar, *Peer-to-Peer Information Retrieval*, UT

2012-30 Alina Pommeranz, *Designing Human-Centered Systems for Reflective Decision Making*, TUD

2012-31 Emily Bagarukayo, *A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure*, RUN

2012-32 Wietske Visser, *Qualitative Multi-Criteria Preference Representation and Reasoning*, TUD

2012-33 Rory Sie, *Coalitions in Cooperation Networks (COCOON)*, OU

2012-34 Pavol Jancura, *Evolutionary Analysis in PPI Networks and Applications*, RUN

2012-35 Evert Haasdijk, *Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics*, VU

2012-36 Denis Ssebugwawo, *Analysis and Evaluation of Collaborative Modeling Processes*, RUN

2012-37 Agnes Nakakawa, *A Collaboration Process for Enterprise Architecture Creation*, RUN

2012-38 Selmar Smit, *Parameter Tuning and Scientific Testing in Evolutionary Algorithms*, VU

2012-39 Hassan Fatemi, *Risk-Aware Design of Value and Coordination Networks*, UT

2012-40 Agus Gunawan, *Information Access for SMEs in Indonesia*, UvT

2012-41 Sebastian Kelle, *Game Design Patterns for Learning*, OU

2012-42 Dominique Verpoorten, *Reflection Amplifiers in Self-Regulated Learning*, OU

2012-43 *Withdrawn*

2012-44 Anna Tordai, *On Combining Alignment Techniques*, VU

2012-45 Benedikt Kratz, *A Model and Language for Business-aware Transactions*, UvT

2012-46 Simon Carter, *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation*, UvA

2012-47 Manos Tsagkias, *Mining Social Media: Tracking Content and Predicting Behavior*, UvA

2012-48 Jorn Bakker, *Handling Abrupt Changes in Evolving Time-series Data*, TUE

2012-49 Michael Kaisers, *Learning Against Learning - Evolutionary Dynamics of Reinforcement Learning Algorithms in Strategic Interactions*, UM

2012-50 Steven van Kervel, *Ontology-driven Enterprise Information Systems Engineering*, TUD

2012-51 Jeroen de Jong, *Heuristics in Dynamic Sceduling; a Practical Framework with a Case Study in Elevator Dispatching*, TUD

—2013—

2013-01 Viorel Milea, *News Analytics for Financial Decision Support*, EUR

2013-02 Erietta Liarou, *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing*, CWI

2013-03 Szymon Klarman, *Reasoning with Contexts in Description Logics*, VU

2013-04 Chetan Yadati, *Coordinating Autonomous Planning and Scheduling*, TUD

2013-05 Dulce Pumareja, *Groupware Requirements Evolutions Patterns*, UT

2013-06 Romulo Goncalves, *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience*, CWI

2013-07 Giel van Lankveld, *Quantifying Individual Player Differences*, UT

2013-08 Robbert-Jan Merk, *Making Enemies: Cognitive Modeling for Opponent Agents in Fighter Pilot Simulators*, VU

2013-09 Fabio Gori, *Metagenomic Data Analysis: Computational Methods and Applications*, RUN

2013-10 Jeewanie Jayasinghe Arachchige, *A Unified Modeling Framework for Service Design*, UvT

2013-11 Evangelos Pournaras, *Multi-level Reconfigurable Self-organization in Overlay Services*, TUD

2013-12 Marian Razavian, *Knowledge-driven Migration to Services*, VU

2013-13 Mohammad Safiri, *Service Tailoring: User-centric Creation of Integrated IT-based Homecare Services to Support Independent Living of Elderly*, UT

2013-14 Jafar Tanha, *Ensemble Approaches to Semi-Supervised Learning Learning*, UvA

2013-15 Daniel Hennes, *Multiagent Learning - Dynamic Games and Applications*, UM

2013-16 Eric Kok, *Exploring the Practical Benefits of Argumentation in Multi-agent Deliberation*, UU

2013-17 Koen Kok, *The PowerMatcher: Smart Coordination for the Smart Electricity Grid*, VU

2013-18 Jeroen Janssens, *Outlier Selection and One-Class Classification*, UvT

2013-19 Renze Steenhuizen, *Coordinated Multi-Agent Planning and Scheduling*, TUD

2013-20 Katja Hofmann, *Fast and Reliable Online Learning to Rank for Information Retrieval*, UvA

2013-21 Sander Wubben, *Text-to-text Generation by Monolingual Machine Translation*, UvT

2013-22 Tom Claassen, *Causal Discovery and Logic*, RUN

2013-23 Patricio de Alencar Silva, *Value Activity Monitoring*, UvT

2013-24 Haitham Bou Ammar, *Automated Transfer in Reinforcement Learning*, UM

2013-25 Agnieszka Anna Latoszek-Berendsen, *Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System*, UM

2013-26 Alireza Zarghami, *Architectural Support for Dynamic Homecare Service Provisioning*, UT

2013-27 Mohammad Huq, *Inference-based Framework Managing Data Provenance*, UT

2013-28 Frans van der Sluis, *When Complexity Becomes Interesting: An Inquiry into the Information eXperience*, UT

2013-29 Iwan de Kok, *Listening Heads*, UT

2013-30 Joyce Nakatumba, *Resource-Aware Business Process Management: Analysis and Support*, TUE

2013-31 Dinh Khoa Nguyen , *Blueprint Model and Language for Engineering Cloud Applications*, UvT

2013-32 Kamakshi Rajagopal, *Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development*, OUN
2013-33 Qi Gao, *User Modeling and Personalization in the Microblogging Sphere*, TUD
2013-34 Kien Tjin-Kam-Jet (UT), *Distributed Deep Web Search*, UT
2013-35 Abdallah El Ali, *Minimal Mobile Human Computer Interaction*, UVA
2013-36 Than Lam Hoang, *Pattern Mining in Data Streams*, TUe
2013-37 Dirk Börner, *Ambient Learning Displays*, OUN
2013-38 Eelco den Heijer, *Autonomous Evolutionary Art*, VU
2013-39 Joop de Jong, *A Method for Enterprise Ontology based Design of Enterprise Information Systems*, TUD
2013-40 Pim Nijssen, *Monte-Carlo Tree Search for Multi-Player Games*, UM
2013-41 Jochem Liem, *Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning*, UVA
2013-42 Léon Planken, *Algorithms for Simple Temporal Reasoning*, TUD
2013-43 Marc Bron, *Exploration and Contextualization through Interaction and Concepts*, UVA

—2014—

2014-01 Nicola Barile, *Studies in Learning Monotone Models from Data*, UU
2014-02 Fiona Tuliyano, *Combining System Dynamics with a Domain Modeling Method*, RUN
2014-03 Sergio Raul Duarte Torres, *Information Retrieval for Children: Search Behavior and Solutions*, UT
2014-04 Hanna Jochmann-Mannak, *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation*, UT
2014-05 Jurriaan van Reijsen, *Knowledge Perspectives on Advancing Dynamic Capability*, UU
2014-06 Damian Tamburri, *Supporting Networked Software Development*, VU
2014-07 Arya Adriansyah, *Aligning Observed and Modeled Behavior*, TUE
2014-08 Samur Araujo, *Data Integration over Distributed and Heterogeneous Data Endpoints*, TUD
2014-09 Philip Jackson, *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language*, UvT
2014-10 Ivan Salvador Razo Zapata, *Service Value Networks*, VU
2014-11 Janneke van der Zwaan, *An Empathic Virtual Buddy for Social Support*, TUD
2014-12 Willem van Willigen, *Look Ma, No Hands: Aspects of Autonomous Vehicle Control*, VU
2014-13 Arlette van Wissen, *Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains*, VU
2014-14 Yangyang Shi, *Language Models With Meta-information*, TUD
2014-15 Natalya Mogles, *Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare*, VU
2014-16 Krystyna Milian, *Supporting trial recruitment and design by automatically interpreting eligibility criteria*, VU
2014-17 Kathrin Dentler VU, *Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability*, VU
2014-18 Mattijs Ghijsen, *Methods and Models for the Design and Study of Dynamic Agent Organizations*, UVA
2014-19 Vinicius Ramos, *Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support*, TUE
2014-20 Mena Habib, *Named Entity Extraction and Disambiguation for Informal Text: The Missing Link*, UT
2014-21 Kassidy Clark, *Negotiation and Monitoring in Open Environments*, TUD
2014-22 Marieke Peeters, *Personalized Educational Games - Developing agent-supported scenario-based training*, UU
2014-23 Eleftherios Sidirourgos , *Space Efficient Indexes for the Big Data Era*, UvA/CWI
2014-24 Davide Ceolin, *Trusting Semi-structured Web Data*, VU
2014-25 **2014-26** Tim Baarslag, *What to Bid and When to Stop*, TUD
2014-27 Rui Jorge Almeida, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty*, EUR
2014-28 Anna Chmielowiec, *Decentralized k-Clique Matching*, VU
2014-29 Jaap Kabbedijk, *Variability in Multi-Tenant Enterprise Software*, UU
2014-30 Peter de Cock, *Anticipating Criminal Behaviour*, UvT
2014-31 Leo van Moergestel, *Agent Technology in Agile Multiparallel Manufacturing and Product Support*, UU
2014-32 Naser Ayat, *On Entity Resolution in Probabilistic Data*, UvA
2014-33 Tesfa Tegegne, *Service Discovery in eHealth*, RUN
2014-34 Christina Manteli, *The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.*, VU
2014-35 Joost van Ooijen, *Cognitive Agents in Virtual Worlds: A Middleware Design Approach*, UU
2014-36 Joos Buijs, *Flexible Evolutionary Algorithms for Mining Structured Process Models*, TUE
2014-37 Maral Dadvar, *Experts and Machines United Against Cyberbullying*, UT
2014-38 Danny Plass-Oude Bos, *Making brain-computer interfaces better: improving usability through post-processing.*, UT
2014-39 Jasmina Maric, *Web Communities, Immigration, and Social Capital*, UvT
2014-40 Walter Omona, *A Framework for Knowledge Management Using ICT in Higher Education*, RUN
2014-41 Frederic Hogenboom, *Automated Detection of Financial Events in News Text*, EUR
2014-42 Carsten Eijckhof, *Contextual Multidimensional Relevance Models*, CWI/TUD
2014-43 Kevin Vlaanderen, *Supporting Process Improvement*

using Method Increments, UU

2014-44 Paulien Meesters, *Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.*, UvT

2014-45 Birgit Schmitz, *Mobile Games for Learning: A Pattern-Based Approach*, OUN

2014-46 Ke Tao, *Social Web Data Analytics: Relevance, Redundancy, Diversity*, TUD

2014-47 Shangsong Liang, *Fusion and Diversification in Information Retrieval*, UVA

—–2015—–

2015-01 Niels Netten, *Machine Learning for Relevance of Information in Crisis Response*, UvA

2015-02 Faiza Bukhsh, *Smart auditing: Innovative Compliance Checking in Customs Controls*, UvT

2015-03 Twan van Laarhoven, *Machine learning for network data*, RUN

2015-04 Howard Spoelstra (OUN), *Collaborations in Open Learning Environments*, OUN

2015-05 Christoph Bösch, *Cryptographically Enforced Search Pattern Hiding*, UT

2015-06 Farideh Heidari, *Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes*, TUD

2015-07 Maria-Hendrike Peetz, *Time-Aware Online Reputation Analysis*, UvA

2015-08 Jie Jiang, *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions*, TUD

2015-09 Randy Klaassen, *HCI Perspectives on Behavior Change Support Systems*, UT

2015-10 Henry Hermans, *OpenU: design of an integrated system to support lifelong learning*, OUN

2015-11 Yongming Luo, *Designing algorithms for big graph datasets: A study of computing bisimulation and joins*, TUE

2015-12 Julie M. Birkholz, *Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks*, VU

2015-13 Giuseppe Procaccianti (VU), *Energy-Efficient Software*, VU

2015-14 Bart van Straalen, *A cognitive approach to modeling bad news conversations*, UT

2015-15 Klaas Andries de Graaf, *Ontology-based Software Architecture Documentation*, VU

2015-16 Changyun Wei, *Cognitive Coordination for Cooperative Multi-Robot Teamwork*, UT

2015-17 André van Cleeff, *Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs*, UT

2015-18 Holger Pirk, *Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories*, CWI

2015-19 Bernardo Tabuenca (OUN), *Ubiquitous Technology for Lifelong Learners*, OUN

2015-20 Lois Vanhée, *Using Culture and Values to Support Flexible Coordination*, UU

2015-21 Sibren Fetter (OUN), *Using Peer-Support to Expand and Stabilize Online Learning*, OUN

2015-22 **2015-23** Luit Gazendam, *Cataloguer Support in Cultural Heritage*, VU

2015-24 Richard Berendsen, *Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation*, UVA

2015-25 Steven Woudenberg, *Bayesian Tools for Early Disease Detection*, UU

2015-26 Alexander Hogenboom, *Sentiment Analysis of Text Guided by Semantics and Structure*, EUR

2015-27 Sándor Héman, *Updating compressed colomn stores*, CWI

2015-28 Janet Bagorogoza, *Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO*, TiU

2015-29 Hendrik Baier, *Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains*, UM

2015-30 Kiavash Bahreini, *Real-time Multimodal Emotion Recognition in E-Learning*, OU

2015-31 Yakup Koç, *On the robustness of Power Grids*, TUD

2015-32 Jerome Gard, *Corporate Venture Management in SMEs*, UL

2015-33 Frederik Schadd, *Ontology Mapping with Auxiliary Resources*, TUD

2015-34 Victor de Graaf, *Gesocial Recommender Systems*, UT

2015-35 Jungxao Xu, *Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction*, TUD

—–2016—–

2016-01 Syed Saiden Abbas, *Recognition of Shapes by Humans and Machines*, RUN

2016-02 Michiel Christiaan Meulendijk, *Optimizing medication reviews through decision support: prescribing a better pill to swallow*, UU

2016-03 Maya Sappelli, *Knowledge Work in Context: User Centered Knowledge Worker Support*, RUN

2016-04 Laurens Rietveld, *Publishing and Consuming Linked Data*, VU

2016-05 Evgeny Sherkhonov, *Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers*, UVA

2016-06 Michel Wilson, *Robust scheduling in an uncertain environment*, TUD

2016-07 Jeroen de Man, *Measuring and modeling negative emotions for virtual training*, VU

2016-08 Matje van de Camp (TiU), *A Link to the Past: Constructing Historical Social Networks from Unstructured Data*, TiU

2016-09 Archana Nottamkandath, *Trusting Crowdsourced Information on Cultural Artefacts*, VU

2016-10 George Karafotias, *Parameter Control for Evolutionary Algorithms*, VUA

2016-11 Anne Schuth, *Search Engines that Learn from Their Users*, UVA

2016-12 Max Knobbout, *Logics for Modelling and Verifying Normative Multi-Agent Systems*, UU

2016-13 Nana Baah Gyan, *The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach*, VU

2016-14 Ravi Khadka, *Revisiting Legacy Software System Modernization*, UU

2016-15 Steffen Michels, *Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments*, RUN

2016-16 Guangliang Li, *Socially Intelligent Autonomous Agents that Learn from Human Reward*, UVA

2016-17 Berend Weel, *Towards Embodied Evolution of Robot Organisms*, VU

2016-18 Albert Meroño Peñuela, *Refining Statistical Data on the Web*, VU

2016-19 Julia Efremova, *Mining Social Structures from Genealogical Data*, TU/e

2016-20 Daan Odijk, *Context & Semantics in News & Web Search*, UVA

2016-21 Alejandro Moreno Célleri, *From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground*, UT

2016-22 Grace Lewis, *Software Architecture Strategies for Cyber-Foraging Systems*, VU

2016-23 Fei Cai, *Query Auto Completion in Information Retrieval*, UVA
2016-24 Brend Wanders, *Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach*, UT
2016-25 Julia Kiseleva, *Using Contextual Information to Understand Searching and Browsing Behavior*, TU/e
2016-26 Dilhan Thilakarathne, *In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains*, VU
2016-27 Wen Li, *Understanding Geo-spatial Information on Social Media*, TUD
2016-28 Mingxin Zhang, *Large-scale Agent-based Social Simulation - A study on epidemic prediction and control*, TUD
2016-29 Nicolas Höning, *Peak reduction in decentralised electricity systems - Markets and prices for flexible planning*, TUD
2016-30 Ruud Mattheij, *The Eyes Have It*, UvT
2016-31 Mohammad Khelghati (UT), *Deep web content monitoring*, UT
2016-32 Eelco Vriezekolk, *Assessing Telecommunication Service Availability Risks for Crisis Organisations*, UT
2016-33 Peter Bloem, *Single Sample Statistics, exercises in learning from just one example*, UVA
2016-34 Dennis Schunselaar, *Configurable Process Trees: Elicitation, Analysis, and Enactment*, TUE
2016-35 Zhaochun Ren, *Monitoring Social Media: Summarization, Classification and Recommendation*, UVA
2016-36 Daphne Karreman, *Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies*, UT
2016-37 Giovanni Sileno (UvA), *Aligning Law and Action - a conceptual and computational inquiry*, UvA
2016-38 Andrea Minuto, *Materials that Matter - Smart Materials meet Art & Interaction Design*, UT
2016-39 Merijn Bruijnes, *Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect*, UT
2016-40 Christian Detweiler, *Accounting for Values in Design*, TUD
2016-41 Thomas King, *Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance*, TUD
2016-42 Spyros Martzoukos, *Combinatorial and Compositional Aspects of Bilingual Aligned Corpora*, UVA
2016-43 Saskia Koldijk, *Context-Aware Support for Stress Self-Management: From Theory to Practice*, RUN
2016-44 Thibault Sellam, *Automatic Assistants for Database Exploration*, UVA
2016-45 Bram van de Laar, *Experiencing Brain-Computer Interface Control*, UT
2016-46 Jorge Gallego Perez, *Robots to Make you Happy*, UT
2016-47 Christina Weber, *Real-time foresight - Preparedness for dynamic innovation networks*, UL
2016-48 Tanja Buttler, *Collecting Lessons Learned*, TUD
2016-49 Gleb Polevoy, *Participation and Interaction in Projects. A Game-Theoretic Analysis*, TUD
2016-50 Yan Wang, *The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains*, UVT

—2017—

2017-01 Jan-Jaap Oerlemans, *Investigating Cybercrime*, UL
2017-02 Sjoerd Timmer, *Designing and Understanding Forensic Bayesian Networks using Argumentation*, UU
2017-03 Daniël Harold Telgen, *Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines*, UU
2017-04 Mrunal Gawade, *Multi-core Parallelism in a Column-store*, CWI
2017-05 Mahdieh Shadi, *Collaboration Behavior*, UVA
2017-06 Damir Vandic, *Intelligent Information Systems for Web Product Search*, EUR
2017-07 Roel Bertens, *Insight in Information: from Abstract to Anomaly*, UU
2017-08 Rob Konijn, *Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detection and Subgroup Discovery*, VU
2017-09 Dong Nguyen, *Text as Social and Cultural Data: A Computational Perspective on Variation in Text*, UT
2017-10 Robby van Delden, *(Steering) Interactive Play Behavior*, UT
2017-11 Florian Kunneman, *Modelling patterns of time and emotion in Twitter #anticipointment*, RUN
2017-12 Sander Leemans, *Robust Process Mining with Guarantees*, TUE
2017-13 Gijs Huisman, *Social Touch Technology - Extending the reach of social touch through haptic technology*, UT
2017-14 Shoshannah Tekofsky, *You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior*, UvT
2017-15 Peter Berck, *Memory-Based Text Correction*, RUN
2017-16 Aleksandr Chuklin, *Understanding and Modeling Users of Modern Search Engines*, UVA
2017-17 Daniel Dimov, *Crowdsourced Online Dispute Resolution*, UL
2017-18 Ridho Reinanda, *Entity Associations for Search*, UVA
2017-19 Jeroen Vuurens, *Proximity of Terms, Texts and Semantic Vectors in Information Retrieval*, UT
2017-20 Mohammadbashir Sedighi, *Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility*, TUD
2017-21 Jeroen Linssen, *Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds)*, UT
2017-22 Sara Magliacane, *Logics for causal inference under uncertainty*, VU
2017-23 David Graus, *Entities of Interest — Discovery in Digital Traces*, UVA
2017-24 Chang Wang (TUD), *Use of Affordances for Efficient Robot Learning*, TUD
2017-25 Veruska Zamborlini, *Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search*, VU
2017-29 Merel Jung, *Socially intelligent robots that understand and respond to human touch*, UT
2017-27 Michiel Joosse, *Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors*, UT
2017-28 John Klein, *Architecture Practices for Complex Contexts*, VU
2017-29 Adel Alhuraibi, *From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT"*, UvT
2017-30 Wilma Latuny, *The Power of Facial Expressions*, UvT
2017-31 Ben Ruijl, *Advances in computational methods for QFT calculations*, UL
2017-32 Thaer Samar, *Access to and Retrievability of Content in Web Archives*, RUN
2017-33 Brigit van Loggem, *Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity*, OU
2017-34 Maren Scheffel, *The Evaluation Framework for Learning Analytics*, OU
2017-35 Martine de Vos, *Interpreting natural science spreadsheets*, VU
2017-36 Yuanhao Guo, *Shape Analysis for Phenotype Charac-*

terisation from High-throughput Imaging, UL
2017-37 Alejandro Montes Garcia , *WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy*, TUE
2017-38 Alex Kayal, *Normative Social Applications*, TUD
2017-39 Sara Ahmadi, *Exploiting properties of the human auditory system and compressive sensing methods to increase noise robustness in ASR*, RUN
2017-40 Altaf Hussain Abro, *Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems*, VUA
2017-41 Adnan Manzoor, *Minding a Healthy Lifestyle: An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle*, VUA
2017-42 Elena Sokolova, *Causal discovery from mixed and missing data with applications on ADHD datasets*, RUN
2017-43 Maaike de Boer, *Semantic Mapping in Video Retrieval*, RUN
2017-44 Garm Lucassen, *Understanding User Stories - Computational Linguistics in Agile Requirements Engineering*, UU
2017-45 Bas Testerink, *Decentralized Runtime Norm Enforcement*, UU
2017-46 Jan Schneider, *Sensor-based Learning Support*, OU
2017-47 Jie Yang, *Crowd Knowledge Creation Acceleration*, TUD
2017-48 Angel Suarez, *Collaborative inquiry-based learning*, OU

—2018—

2018-01 Han van der Aa, *Comparing and Aligning Process Representations*, VUA
2018-02 Felix Mannhardt, *Multi-perspective Process Mining*, TUE
2018-03 Steven Bosems, *Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction*, UT
2018-04 Jordan Janeiro, *Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks*, TUD
2018-05 Hugo Huurdeman, *Supporting the Complex Dynamics of the Information Seeking Process*, UVA
2018-06 Dan Ionita, *Model-Driven Information Security Risk Assessment of Socio-Technical Systems*, UT
2018-07 Jieting Luo, *A formal account of opportunism in multi-agent systems*, UU
2018-08 Rick Smetsers, *Advances in Model Learning for Software Systems*, RUN
2018-09 Xu Xie, *Data Assimilation in Discrete Event Simulations*, TUD
2018-10 Julienka Mollee, *Moving forward: supporting physical activity behavior change through intelligent technology*, VUA
2018-11 Mahdi Sargolzaei, *Enabling Framework for Service-oriented Collaborative Networks*, UVA
2018-12 Xixi Lu, *Using behavioral context in process mining*, TUE
2018-13 Seyed Amin Tabatabaei, *Computing a Sustainable Future*, VUA
2018-14 Bart Joosten, *Detecting Social Signals with Spatiotemporal Gabor Filters*, UVT
2018-15 Naser Davarzani, *Biomarker discovery in heart failure*, UM
2018-16 Jaebok Kim, *Automatic recognition of engagement and emotion in a group of children*, UT
2018-17 Jianpeng Zhang, *On Graph Sample Clustering*, TUE
2018-18 Henriette Nakad, *De Notaris en Private Rechtspraak*, UL
2018-19 Minh Duc Pham, *Emergent relational schemas for RDF*, VUA
2018-20 Manxia Liu, *Time and Bayesian Networks*, RUN
2018-21 Aad Slootmaker, *EMERGO: a generic platform for authoring and playing scenario-based serious games*, OUN
2018-22 Eric Fernandes de Mello Araujo, *Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks*, VUA
2018-23 Kim Schouten, *Semantics-driven Aspect-Based Sentiment Analysis*, EUR
2018-24 Jered Vroon, *Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots*, UT
2018-25 Riste Gligorov, *Serious Games in Audio-Visual Collections*, VUA
2018-26 Roelof Anne Jelle de Vries, *Theory-Based and Tailor-Made: Motivational Messages for Behavior Change Technology*, UT
2018-27 Maikel Leemans, *Hierarchical Process Mining for Scalable Software Analysis*, TUE
2018-28 Christian Willemse, *Social Touch Technologies: How they feel and how they make you feel*, UT
2018-29 Yu Gu, *Emotion Recognition from Mandarin Speech*, UVT
2018-30 Wouter Beek, *The "K" in "semantic web" stands for "knowledge": scaling semantics to the web*

—2019—

2019-01 Rob van Eijk, *Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification*, UL
2019-02 Emmanuelle Beauxis Aussalet, *Statistics and Visualizations for Assessing Class Size Uncertainty*, CWI, UU
2019-31 Eduardo Gonzalez Lopez de Murillas, *Process Mining on Databases: Extracting Event Data from Real Life Data Sources*, TUE
2019-04 Ridho Rahmadi, *Finding stable causal structures from clinical data*, RUN
2019-05 Sebastiaan van Zelst, *Process Mining with Streaming Data*, TUE
2019-06 Chris Dijkshoorn, *Nichesourcing for Improving Access to Linked Cultural Heritage Datasets*, VU
2019-07 Soude Fazel, *Recommender Systems in Social Learning Platforms*, TUD
2019-08 Frits de Nijs, *Resource-constrained Multi-agent Markov Decision Processes*, TUD
2019-09 Fahimeh Alizadeh Moghaddam, *Self-adaptation for energy efficiency in software systems*, UVA
2019-10 Qing Chuan Ye, *Multi-objective Optimization Methods for Allocation and Prediction*, EUR
2019-11 Yue Zhao, *Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs*, TUD
2019-12 Jacqueline Heinerman, *Better Together*, VU
2019-13 Guanliang Chen, *MOOC Analytics: Learner Modeling and Content Generation*, TUD
2019-14 Daniel Davis, *Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses*, TUD
2019-15 Erwin Walraven, *Planning under Uncertainty in Constrained and Partially Observable Environments*, TUD

2019-16 Guangming Li, *Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models*, TUE
2019-17 Ali Hurriyetoglu, *Extracting actionable information from microtexts*, RUN
2019-18 Gerard Wagenaar, *Artefacts in Agile Team Communication*, UU

2019-19 Vincent Koeman, *Tools for Developing Cognitive Agents*, TUD
2019-20 Chide Groenouwe, *Fostering technically augmented human collective intelligence*, UU
2019-21 Cong Liu, *Software Data Analytics: Architectural Model Discovery and Design Pattern Detection*, TUE
2019-22 Martin van den Berg, *Improving IT Decisions with Enterprise Architecture*, VU
2019-23 Qin Liu, *Intelligent Control Systems: Learning, Interpreting, Verification*, TUD
2019-24 Anca Dumitrache, *Truth in Disagreement - Crowdsourcing Labeled Data for Natural Language Processing*, VU
2019-25 Emiel van Miltenburg, *Pragmatic factors in (automatic) image description*, VU
2019-26 Prince Singh, *An Integration Platform for Synchromodal Transport*, UT
2019-27 Alessandra Antonaci, *The Gamification Design Process applied to (Massive) Open Online Courses*, OUN
2019-28 Esther Kuindersma, *Cleared for take-off: Game-based learning to prepare airline pilots for critical situations*, UL
2019-29 Daniel Formolo, *Using virtual agents for simulation and training of social skills in safety-critical circumstances*, VU
2019-30 Vahid Yazdanpanah, *Multiagent Industrial Symbiosis Systems*, UT
2019-31 Milan Jelisavcic, *Alive and Kicking: Baby Steps in Robotics*, VU
2019-32 Chiara Sironi, *Monte-Carlo Tree Search for Artificial General Intelligence in Games*, UM
2019-33 Anil Yaman, *Evolution of Biologically Inspired Learning in Artificial Neural Networks*, TUE
2019-34 Negar Ahmadi, *EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES*, TUE
2019-35 Lisa Facey-Shaw, *Gamification with digital badges in learning programming*, OUN
2019-36 Kevin Ackermans, *Designing Video-Enhanced Rubrics to Master Complex Skills*, OUN
2019-37 Jian Fang, *Database Acceleration on FPGAs*, TUD
2019-38 Akos Kadar, *Learning visually grounded and multilingual representations*, OUN

—2020—

2020-01 Armon Toubman, *Calculated Moves: Generating Air Combat Behaviour*, UL
2020-02 Marcos de Paula Bueno, *Unraveling Temporal Processes using Probabilistic Graphical Models*, UL
2020-03 Mostafa Deghani, *Learning with Imperfect Supervision for Language Understanding*, UvA
2020-04 Maarten van Gompel, *Context as Linguistic Bridges*, RUN
2020-05 Yulong Pei, *On local and global structure mining*, TUE
2020-06 Preethu Rose Anish, *Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support*, UT
2020-07 Wim van der Vegt, *Towards a software architecture for reusable game components*, OUN

2020-08 Ali Mirsoleimani, *Structured Parallel Programming for Monte Carlo Tree Search*, UL

2020-09 Myriam Traub, *Measuring Tool Bias and Improving Data Quality for Digital Humanities Research*, UU
2020-10 Alifah Syamsiyah, *In-database Preprocessing for Process Mining*, TU/e

2020-11 Sepideh Mesbah, *Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models*, TUD

2020-12 Ward van Breda, *Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment*, VU

2020-13 Marco Virgolin, *Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming*, CWI

2020-14 Mark Raasveldt, *Integrating Analytics with Relational Databases*, CWI/UL

2020-15 Konstantinos Georgiadis , *Smart CAT: Machine Learning for Configurable Assessments in Serious Games*, OUN

2020-16 Ilona Wilmont, *Cognitive Aspects of Conceptual Modelling*, RUN

2020-17 Daniele Di Mitri , *The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences*, OUN

2020-18 Georgios Methenitis, *Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems*, TUD

2020-19 Guido van Capelleveen, *Industrial Symbiosis Recommender Systems*, UT

2020-20 Albert Hankel, *Embedding Green ICT Maturity in Organisations*, VU
2020-21 Karine da Silva Miras de Araujo, *Where is the robot?: Life as it could be*, VU
2020-22 Maryam Masoud Khamis, *Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar*, RUN
2020-23 Rianne Conijn, *The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging*, UT
2020-24 Lenin da Nobrega Medeiros , *How are you feeling, human? Towards emotionally supportive chatbots*, VUA/RUN
2020-25 Xin Du, *The Uncertainty in Exceptional Model Mining*, TUE
2020-26 Krzysztof Leszek Sadowski, *GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization*, UU
2020-27 Ekaterina Muravyeva, *Personal data and informed consent in an educational context*, TUD
2020-28 Bibeg Limbu, *Multimodal interaction for deliberate practice: Training complex skills with augmented reality*, TUD
2020-29 Ioan Gabriel Bucur, *Being Bayesian about Causal Inference*, RUN
2020-30 Bob Zadok Blok, *Creatief, Creatieve, Creatiefst* UL
2020-31 Gongjin Lan, *Learning better – From Baby to Better*, VU
2020-32 Jason Rhuggenaath, *Revenue management in online markets: pricing and online advertising*, TUE
2020-33 Rick Gilsing, *Supporting service-dominant business model evaluation in the context of business model innovation*, TUE
2020-34 Anna Bon, *Intervention or Collaboration? Redesigning Information and Communication Technologies for Development*, MU
2020-35 Siamak Farshidi, *Multi-Criteria Decision-Making in*

Software Production, UU

—–2021—–

2021-01 Francisco Xavier Dos Santos Fonseca , *Location-based Games for Social Interaction in Public Space*, TUD
2021-02 Rijk Mercuur, *Simulating Human Routines:Integrating Social Practice Theory in Agent-Based Models*, TUD
2021-03 Seyyed Hadi Hashemi, *Modeling Users Interacting with Smart Devices*, UVA
2021-04 Ioana Jivet, *The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning*, OU
2021-05 Davide Dell'Anna, *Data-Driven Supervision of Autonomous Systems*, UU
2021-06 Daniel Davison, *"Hey robot, what do you think?" How children learn with a social robot*, UT
2021-07 Armel Lefebvre, *Research data management for open science*, UU
2021-08 Nardie Fanchamps, *The Influence of Sense-Reason-Act Programming on Computational Thinking*, OU
2021-09 Cristina Zaga, *The Design of Robothings. Non-Anthropomorphic and Non-Verbal Robots to Promote Children's Collaboration Through Play*, UT
2021-10 Quinten Meertens, *Misclassification Bias in Statistical Learning*, UvA
2021-11 Anne van Rossum, *Nonparametric Bayesian Methods in Robotic Vision*, UL
2021-12 Lei Pi, *External Knowledge Absorption in Chinese SMEs*, UL
2021-13 Bob R. Schadenberg, *Robots for Autistic Children: Understanding and Facilitating Predictability for Engagement in Learning*, UT
2021-14 Negin Samaeemofrad, *Business Incubators: The Impact of Their Support*, UL
2021-15 Onat Ege Adali, *Transformation of Value Propositions into Resource Re-Configurations through the Business Services Paradigm*, TU/e
2021-16 Esam A. H. Ghaleb, *BIMODAL EMOTION RECOGNITION FROM AUDIO-VISUAL CUES*, UM
2021-17 Dario Dotti, *Human Behavior Understanding from motion and bodily cues using deep neural networks*, UM
2021-8 Remi Wieten, *Bridging the Gap Between Informal Sense-Making Tools and Formal Systems - Facilitating the Construction of Bayesian Networks and Argumentation Frameworks*, UU
2021-19 Roberto Verdecchia, *Architectural Technical Debt: Identification and Management*, VU
2021-20 Masoud Mansoury, *Understanding and Mitigating Multi-Sided Exposure Bias in Recommender Systems*, TU/e
2021-21 Pedro Thiago Timbó Holanda, *Progressive Indexes*, CWI
2021-22 Sihang Qiu, *Conversational Crowdsourcing*, TUD
2021-23 Hugo Manuel Proença, *Robust rules for prediction and description*, LIACS
2021-24 Kaijie Zhu, *On Efficient Temporal Subgraph Query Processing*, TUE
2021-25 Eoin Martino Grua, *The Future of E-Health is Mobile: Combining AI and Self-Adaptation to Create Adaptive E-Health Mobile Applications*, VUA
2021-26 Benno Kruit, *Reading the Grid: Extending Knowledge Bases from Human-readable Tables*, CWI & VUA
2021-27 Jelte van Waterschoot, *Personalized and Personal Conversations: Designing Agents Who Want to Connect With You*, UT
2021-28 Christoph Selig, *Understanding the Heterogeneity of Corporate Entrepreneurship Programs*, UL

—–2022—–

2022-1 Judith van Stegeren), *Flavor text generation for role-playing video games*, UT
2022-2 Paulo da Costa, *Data-driven Prognostics and Logistics Optimisation: A Deep Learning Journey*, TU/e
2022-3 Ali el Hassouni, *A Model A Day Keeps The Doctor Away: Reinforcement Learning For Personalized Healthcare*, VUA
2022-4 Ünal Aksu, *A Cross-Organizational Process Mining Framework*, UU
2022-5 Shiwei Liu, *Sparse Neural Network Training with In-Time Over-Parameterization*, TU/e
2022-6 Reza Refaei Afshar, *Machine Learning for Ad Publishers in Real Time Bidding*, TU/e
2022-7 Sambit Praharaj, *Measuring the Unmeasurable? Towards Automatic Co-located Collaboration Analytics*, OU
2022-8 Maikel L. van Eck, *Process Mining for Smart Product Design*, TU/e

TU Delft
DM 2022