Department of Media, Aalto Media Lab

Explorable Information Spaces

Designing Entity Affordances for Fluid Information Exploration

Khalil Klouche

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 192/2019

Explorable Information Spaces

Designing Entity Affordances for Fluid Information Exploration

Khalil Klouche

**Aalto University
School of Arts, Design and Architecture
Department of Media, Aalto Media Lab**

Supervising professor
Lily Diaz-Kommonen, Aalto University, Finland

Thesis advisors
Giulio Jacucci, University of Helsinki, Finland
Lily Diaz-Kommonen, Aalto University, Finland
Tuukka Ruotsalo, University of Helsinki, Finland

Preliminary examiners
Sheelagh Carpendale, Simon Fraser University
Marian Dörk, FH Potsdam

Opponent
Andruid Kerne, Texas A&M University

Aalto University publication series
DOCTORAL DISSERTATIONS 192/2019

ISBN 978-952-60-8774-0 (printed)
ISBN 978-952-60-8775-7 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-8775-7

Unigrafia Oy
Helsinki 2019

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

Aalto University

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Abstract

Author
Khalil Klouche

Name of the doctoral dissertation
Explorable Information Spaces

Publisher School of Arts, Design and Architecture

Unit Department of Media, Aalto Media Lab

Series Aalto University publication series DOCTORAL DISSERTATIONS 192/2019

Field of research Human-Computer Interaction

Date of the defence 5 December 2019

Language English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

The research presented in this dissertation explores interactive techniques designed to overcome limitations of current conventional search tools as primary access points to information, and better support for a wider range of information-seeking behaviors, including exploratory search, serendipity, and orientation. I summarize that goal as *Making information explorable* and define explorability as the quality of physical space that enables humans to become acquainted with it through movement and exploration.

An explorable information space implies situated information, which enables orientation: Choosing a direction instead of formulating queries; Meaningful overviews instead of narrow looks; Persistent spaces allowing growing familiarity, sense-making, and collaboration, instead of quick disposable search sessions.

The promise of an information space with such properties is carried by the notion of entity-oriented information. I address the state of explorability of the information space through the evaluation of entity affordances and visualization techniques. To that end, I demarcate three properties of explorability, i.e., Direction, Orientation, and Continuity, which I use as design drivers in the development of various prototypes and user experiments.

The research process has yielded eight publications, including seven design cases with user experiments, and a position paper. The summarized work consists of an extensive design exploration of the topic at hand, and proposes a variety of interaction techniques that have shown to support information exploration, and together demarcate an alternative paradigm for future information practices.

Keywords Information Exploration, Interface Design

ISBN (printed) 978-952-60-8774-0	**ISBN (pdf)** 978-952-60-8775-7
ISSN (printed) 1799-4934	**ISSN (pdf)** 1799-4942
Location of publisher Helsinki	**Location of printing** Helsinki **Year** 2019
Pages 260	**urn** http://urn.fi/URN:ISBN:978-952-60-8775-7

Acknowledgements

I dedicate the present thesis to the late Dr. Luana Micallef, whose passion, knowledge and availability have helped me countless times.

The research presented in this thesis has been made possible thanks to the infrastructure provided by the Department of Computer Science at University of Helsinki, the Helsinki Institute for Information Technology (HIIT) both at Aalto University and University of Helsinki, and the Media Lab at the Aalto School of Arts, Design and Architecture.

First, I must acknowledge the support of my main advisor and mentor, Professor Giulio Jacucci, who provided me with the best conditions to conduct this research. Giulio's insights, advice, and trust allowed me to pursue ambitious goals and gave me the chance to collaborate with exceptional researchers. I share the credit of this work with him.

I extend my warmest thanks to my supervisor and advisor, Professor Lily Diaz-Kommonen whose teaching helped me grow throughout the process. I am especially grateful for the crucial support Lily provided in the final stage of this research. Her availability and advice have been pivotal in my ability to complete the present thesis.

I would like to thank my advisor Dr. Tuukka Ruotsalo who taught this designer how to be a scientist and provided me with state-of-the-art technology to build amazing prototypes.

I am indebted to all the co-authors of the articles that have emanated from this research: Diogo Cabral, Salvatore Andolina, Valeria Orso, Tung Vuong, Dorota Głowacka, Arto Klami, Pedram Daee, Mats Sjöberg, Mohammad Hoque, Jaakko Peltonen, Hendrik Schneider, Samuel Kaski, Patrik Floréen and Luciano Gamberini. I am especially grateful to Salvatore Andolina as we have worked closely on many research and teaching ventures.

I extend my thanks to my colleagues from the Ubiquitous Interaction Group and HCI: Baris Serim, Imtiaj Ahmed, Kumaripaba Athukorala, Marco Filetti, Ville Harjunen, Antti Jylhä, Chen He, Yi-Ta Hsieh, Matti Nelimarkka, Antti Salovaara, and Andrea Vianello. Special thanks go to Oswald Barral and Andrea Bellucci, who have become close friends.

I thank all the friendly staff at the Department of Computer Science at University of Helsinki, the Aalto School of Arts, Design and Architecture, and the Aalto Media Lab for creating a pleasant work environment.

I am grateful to my friends and family for infusing me with the confidence to pursue my goals, especially my sister Leïla for her active help in the later stages of the writing. Final thanks go to Louise Roy, my best friend and love of my life, for her unwavering support and understanding.

Lausanne, September 2019
Khalil Klouche

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. Designing for Exploratory Search on Touch Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, (CHI '15). New York, NY, USA, pp. 4189-4198, ACM. DOI: 10.1145/2702123.2702489, April 2015.

II Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, (IUI '15). ACM, New York, NY, USA, pp. 300-305, 2015. DOI: 10.1145/2678025.2701401, June 2015.

III Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, (CHIIR '17). ACM, New York, NY, USA, pp. 57-66, 2017. DOI: 10.1145/3020165.3020174, March 2017.

IV Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. QueryTogether: Enabling Entity-Centric Exploration in Multi-Device Collaborative Search. *Information Processing & Manage-*

ment, Elsevier, Volume 54, Issue 6, 2018, pp. 1182-1202,ISSN 0306-4573, DOI: 10.1016/j.ipm.2018.04.005, May 2018.

V Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, (C&C '15). ACM, New York, NY, USA, 103-106, 2015. DOI: 10.1145/2757226.2757252, June 2015.

VI Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Conference on Designing Interactive Systems*, (DIS '18). ACM, New York, NY, USA, 1141-1152, 2018. DOI: 10.1145/3064663.3064728, June 2018.

VII Salvatore Andolina, Pedram Daee, Tung Vuong, Tuukka Ruotsalo, Khalil Klouche, Mats Sjöberg, Samuel Kaski,and Giulio Jacucci. Proactive Recommendation in Context: From Relevant Items to Actionable Entities. Submitted to *ACM Transactions on Computer-Human Interaction*, February 2019.

VIII Khalil Klouche, Tuukka Ruotsalo, and Giulio Jacucci. From Hyperlinks to Hypercues: Entity-Based Affordances for Fluid Information Exploration. In *Proceedings of the 2018 Conference on Designing Interactive Systems*, (DIS '18). ACM, New York, NY, USA, 1141-1152, 2018. DOI: 10.1145/3064663.3064728, June 2018.

Author's Contribution

Publication I: "Designing for Exploratory Search on Touch Devices"

The author was the main contributor of the overall publication, as well as the user interface design and implementation of the studied system. He led collaborative efforts regarding the study design, data analysis and result discussion.

Publication II: "IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search"

The author was the main contributor of the user interface design and implementation of the studied system. He assisted with the study design and result discussion.

Publication III: "Visual Re-Ranking for Multi-Aspect Information Retrieval"

The author was the main contributor of the overall publication, the user interface design and implementation of the studied system, the study design and data analysis.

Publication IV: "QueryTogether: Enabling Entity-Centric Exploration in Multi-Device Collaborative Search"

The author was the main contributor of the user interface design and implementation of the studied system, and assisted with the study design and result discussion.

Publication V: "InspirationWall: Supporting Idea Generation Through Automatic Information Exploration"

The author was the main contributor of the user interface design and implementation of the studied system, and assisted with the study design and result discussion.

Publication VI: "Investigating Proactive Search Support in Conversations"

The author was the main contributor of the user interface design of the studied system.

Publication VII: "Proactive Recommendation in Context: From Relevant Items to Actionable Entities"

The author was the main contributor of the user interface design and implementation of the studied system.

Publication VIII: "From Hyperlinks to Hypercues: Entity-Based Affordances for Fluid Information Exploration"

The author was the main contributor of the overall publication.

Contents

1. Introduction

The research presented in this dissertation explores interactive techniques designed to overcome the limitations of current conventional search tools as primary access points to information, and to support a wider range of information-seeking behaviors, including exploratory search, serendipity and orientation. This first chapter introduces the research context, establishes the objective and scope of the dissertation and the research questions. It then provides an overview of the research contributions from eight publications that originated from the research process. Finally it previews the structure of this work.

1.1 Research Context

The rise of ubiquitous connectivity has dramatically changed our environment, providing access to ever-growing amounts of information, shifting our economy and giving birth to whole new industries. The web is no longer comparable to a very large library but has become the primary place of growth, significance, and struggle in our culture and society [Dörk et al., 2011]. And in that environment, search engines have become our *de facto* point of access and way of finding, filtering and discovering information. Their centrality in our information practices and pervasiveness in our daily lives makes them infrastructure-like, i.e. as important as they are invisible [König and Rasch, 2014]. For these reasons, and while we must acknowledge their usefulness, it is essential to also address their limitations, which are both ethical and technical.

The first ethical limitation pertains to their black-box status. Search engines do not reveal the way they function. We rely on their artificial and arbitrary sense of relevance to sample and rank information of interest, and ultimately be given access to it, with no visibility regarding what has

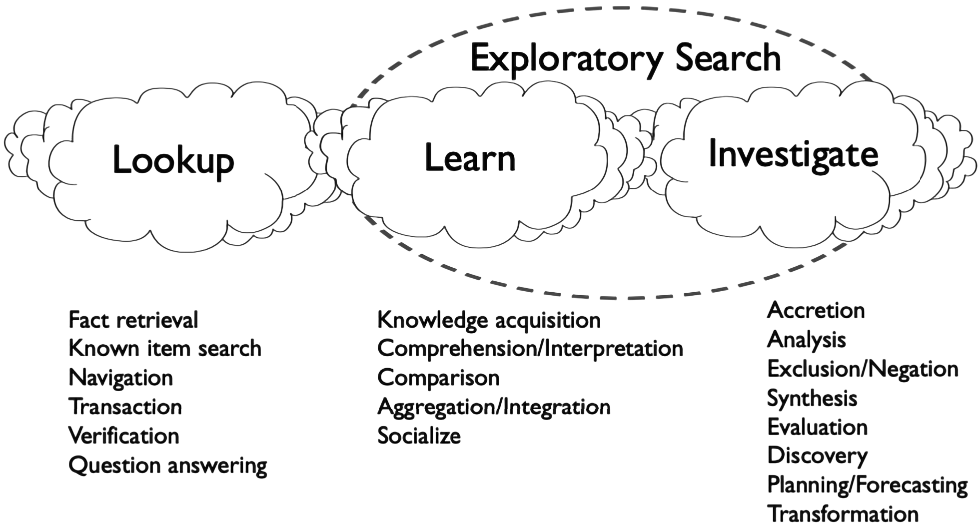


Figure 1.1. Categorization of search activities falling under lookup and exploratory search [Marchionini, 2006].

been left out and no further explanation. The second ethical limitation has to do with the documented impossibility, in spite of their claims, for search engines to be neutral. The simple fact that they are not a public service, but unregulated commercial operations in charge of filtering and ranking information introduces a bias [Lewandowski, 2015]. Furthermore, results to a given query will differ from one user to another as the consequence of some personalization process, preventing any form of objectivity. The third ethical limitation lies in the discrepancy between the perceived simplicity of use, and the actual level of skills required both in formulating effective queries and analysis of returned search results [König and Rasch, 2014]. This results in inequalities regarding the quality of available information, making users potentially vulnerable to filter bubbles [Pariser, 2011] and various interests.

The core issue regarding the three limitations above lie in the lack of control we have over the search engines, and the passive role in which the user is encouraged, which verifies in the type of search behavior fostered by search engines. Marchionini describes fundamental differences between simple search tasks, or *Lookup*, e.g., fact retrieval or known-item search, and more complex tasks, described as *Exploratory Search* [Marchionini, 2006]. Exploratory search tasks differ from lookup in that users engage in them without having a predetermined goal in mind, for example when looking for inspiration or wanting to know more about a given topic. Such tasks, having no predetermined end, become dynamic and potentially long term, implying information needs and strategies that shift and evolve as users learn, discover new information and become familiar with the information space[1] [Capra et al., 2007].

As it is shown in Figure 1.1, exploratory search is not to be reduced to a single type of search activity, but encompasses a wide array of information related activities.

Neglecting ethical concerns mentioned above, current search engines offer an incredibly convenient support for lookup tasks, answering all of our questions in a few milliseconds. However, the lack of support for more complex search tasks contributes to maintaining users in a passive consumer role instead of rewarding active informational behavior. I describe these shortcomings as the following technical limitations:

First, formulating and refining textual queries is known to be difficult. Exploration has users go into information areas with which they are unfamiliar. A user-defined query being built upon pre-acquired knowledge, it creates little opportunity for discovery [White and Roth, 2009]. Furthermore, as the open-ended nature of exploration makes querying an iterative process, that difficulty is made all the more salient.

Second, conventional result lists offer limited support in understanding the related information space. Not only are a few search results, commonly referred to as *ten blue links*, too narrow an access point to offer any sensible overview of available material related to any given query, but nothing tells the user how one result relates to another, or how the ten most relevant results are representative of the information space [Balasubramanian and Cucerzan, 2010]. In other words, are these results redundant, or do they offer complementary directions with respect to the topic at hand? For now, the underlying structure of the data remains hidden, while only an evaluation of individual results can provide that kind of insights.

[1]Throughout this dissertation, I use *Information Space* to describe a set of information objects. Through the spatial aspect, I consider the numerous relationships that potentially link these objects. Among possible relationships, *semantic relatedness* provides a convenient metric distance between objects. Guex provides affordable preliminary definitions of graph theory, which tells us a graph is qualified as *spatial* when its nodes exist in a metric space, i.e., there is a distance between them [Guex, 2016]. The consideration of information as a space is an intrinsic consequence of applying the notion of exploration to information, which suggests paths between points to be uncovered

Third, search is considered and supported as an ephemeral single-user activity. As a result, there is very limited support for keeping track of encountered information or to take advantage of collaborative work. Such limitations make information exploration a dissociated experience, with many missed opportunities to support the user and help her link and make sense[2] of new information.

I advocate instead for greater user awareness and control as well as increased responsibility for what information is encountered. The research presented in this dissertation explores interactive techniques designed to overcome the above-mentioned limitations, and demarcate an alternative paradigm for information practices. I summarize that goal as *making information explorable*, and define explorability as the quality of physical space that enables humans to become acquainted with it through movement and exploration. An explorable information space implies situated information, which enables orientation:

- Choosing a direction instead of formulating queries.
- Meaningful overviews instead of narrow looks.
- Persistent spaces allowing growing familiarity, sense-making, and collaboration instead of quick disposable search sessions.

The promise of an information space with such properties is carried by the notion of *entity*. In information science fields, entities are data elements and objects of interest [Ware, 2012], and constitute references to real-world objects or concepts (e.g., persons, places, movies, topics, and products) [Miliaraki et al., 2015]. "Tom Hanks (actor)", who plays in "Forrest Gump (movie)" can both be entities, as well as "the entire cast of Forrest Gump". Entities offer a flexible way of structuring concepts in a way that is relevant to a context or the task at hand. Entities are structured through their relationships. For example, the entity "Tom Hanks (actor)" and the entity

[2]Sense-making is a largely interdisciplinary concept that can be defined as the process through which people give meaning to their experience [Klein et al., 2006]. In the context of this dissertation, I consider the definition of Russel and colleagues, who define sense-making as *the process of searching for a representation and encoding data in that representation to answer task-specific questions* [Russell et al., 1993]. Through several empirical studies, they were able to provide an operational description of sense-making as: (1) A retrospective analysis of events, (2) Guiding information exploration, (3) A social activity fostering the finding of a common ground, (4) An open-ended process that will consume any amount of invested time resource. Sense-making is not to be confused with mental modeling, i.e. a memory representation of linked concepts and principles, or with situation awareness, i.e. a state of knowledge of current data elements, allowing inferences, predictions, and decision making.

"*Forrest Gump* (movie)" are linked via a relationship of the type "stars in". Entity-linking relationships are numerous. They can be causal or temporal [Ware, 2012]. Together, they form a graph in which entities are nodes and relationships are the edges[3]. While conventional databases store data as lists of items described by a standardized set of features, entity graphs (also known as knowledge graphs or knowledge bases) define each item as a collection of directed links (or typed relationship) to other items. Such data structure allows algorithms to solve complex queries. For example: "Who is the inventor of the paperclip?" can be answered by identifying *paperclip* as an entity, and *inventor of* as a type of relationship. Given a sufficiently comprehensive graph of entities, the answer lies at the end of the identified path).

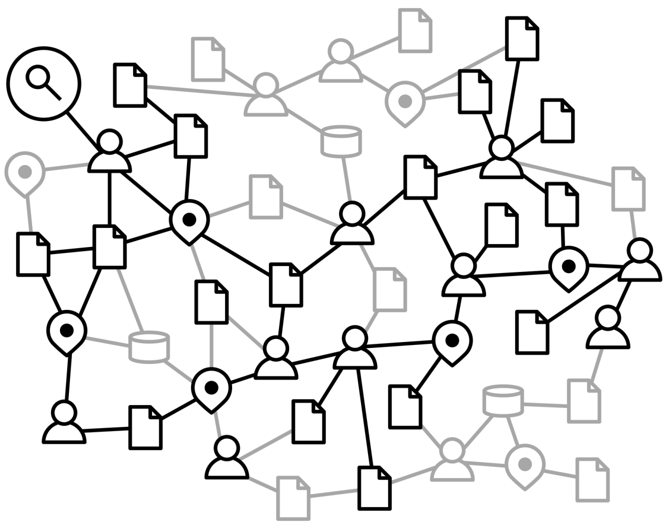


Figure 1.2. The notion of information space is often implicit. Semantic adjacencies exist at a conceptual level, e.g., through the relatedness of the topics discussed. Once available information is meaningfully structured, these conceptual links become explicit and can be made visible and interactive. The information space becomes like a multidimensional medium.

Entity search is investigated by both *Information Retrieval* and *Semantic Web* research communities. In Information Retrieval, entity search is done primarily through statistical methods. Such methods identify entities that coexist within a common content, e.g., an article, through which it infers whether and how they are related. Similar methods allow identifying of entities that are relevant to a given query [Balog, 2018]. Information

[3]Such graphs are commonly used in the Digital Humanities, for example as a means of analysis of a character network in fictional work [Rochat and Triclot, 2017].

Retrieval methods already yield results and are successfully integrated in commonly used search algorithms to improve search results.

Within the Semantic Web community, entity search is mainly tackled as an information structure and path finding problem. Such approach implies creation of a global entity database, or knowledge base, to which point entities referenced in a content [Berners-Lee et al., 2001]. The goal of the Semantic Web, infusing common content on the web with meaning that is simultaneously legible by humans and machines, is ambitious and difficult, and not yet realized [Balog et al., 2010], but it promises plenty of opportunities, not only in the way complex textual queries can be solved, but in how we will interact with information.

Once available information is meaningfully structured, the information space becomes a multidimensional medium ready to be unfolded as explorers pull its threads in various directions and discover content according to their needs, inspiration or chance (Figure 1.2). Utilizing entities interactively for information exploration requires substantial thinking about their affordances[4] and visualization techniques, to come up with fundamental principles that are generalizable to, and across, tasks and search contexts, e.g., academic literature, social media or movies.

1.2 Objective and Scope

The main challenge addressed by this research concerns the state of explorability of the information space, which I intend to address through the development of entity-oriented interaction and visualization techniques.

The goal of enabling explorability in the information space consists of turning a highly abstract activity into an embodied and situated experience. In this context, Suchman's parallel of Human-Machine Interactions as a navigation issue [Suchman, 2007] becomes quite literal, and her model describing the process of meaningful human actions provides us with a useful way to structure the present research in how to support them. Therefore, exploring the design space[5] of entity affordances supporting

[4] An *affordance* is a possibility for action enabled by an object. A ball affords being picked up or thrown. A button affords being pushed, a hyperlink, being clicked. In the case of physical objects, most affordances are visible, as they are constrained by their physical properties, e.g. a small or visibly light object can usually be picked up, an empty container can be filled [Norman, 2013]. In the digital realm, the disconnection between perceived properties of an object and its utility, makes it a central challenge in user interface design [Abras et al., 2004].

[5] *Design space* refers to at least three completely different concepts

information exploration required addressing the following:

1. Enabling the technical means for action, i.e., providing a direction to the exploration. I refer to the resulting property of explorability as *Direction*.

2. Enabling the means to situate such actions allowing for planning and accounting. I refer to the resulting property as pertaining to *Orientation*.

3. Establishing the spatial consistency required for 1. and 2., i.e., bridging gaps across heterogeneous information sources and exploration environments. I refer to the resulting property as pertaining to *Continuity*.

I use these three properties of explorability as design drivers in our various attempts, which together provide a comprehensive approach to the research goal. The following expands on their application.

1. Direction
Search in the information space is a common process through which elements are made retrievable or accessible. In lookup tasks, as the user's intent is clearly defined, search is aimed at an endpoint, e.g. an answer to a question. In exploratory settings, users' intents are less clearly defined and more complex, thus search becomes the process of providing a direction to the exploration. Techniques involving interaction with entities have been investigated [Miliaraki et al., 2015] but are very limited in addressing a user's search intent, especially when compared with typed queries in conventional search engines. Therefore to address this property of explorability, I needed to investigate the design space of entity-based affordances for search, with the objective of developing and evaluating techniques that would yield search results of a quality at least comparable to conventional search methods, for effective use in exploratory settings.

2. Orientation
Orientation refers to the location of something in relation to its surroundings. In terms of ability, it refers to the prerequisites for navigation, or one's capability to situate oneself with respects to one's origin or past locations, therefore enabling decision for future directions. The abstraction of information as a space makes that notion somewhat fuzzy but can be described

---

[Sanders and Westerlund, 2011]. In the context of this dissertation, I define it as: *The imaginary set of all possible solutions to a sub-constrained problem.*

as having a sense of overview and direction [Dörk et al., 2011]. Orientation is simply not supported in commonly available ways of accessing information. Conventional search engines respond to users' queries with a short selection of relevant items with no further indication on how one result relates to others or to the larger context, providing potentially redundant coverage of the topic at hand [Balasubramanian and Cucerzan, 2010]. While interactive visualization methods aimed at supporting sense-making exist and have been investigated [Stasko et al., 2008, Chau et al., 2011], they primarily target professionally trained analysts and are explicitly designed to support problem-solving through an additive approach of providing numerous views and panels to complement each other and achieve an exhaustive understanding of a complex problem.

Addressing support for orientation in the context of this research requires instead developing techniques for visualizing information that enable a user to grow familiar with information areas of her choice, fostering insights but also serendipity, through visualization and manipulation of entities.

3. Continuity

Continuity ensures that all parts of space communicate in a consistent way. Information exploration is potentially long-term, collaborative, and often relies on heterogeneous sources for insights and learning [White and Roth, 2009], including active search, serendipitous finds online and offline, conversations, etc. However, the lack of consistent and direct communication between various sources of information results in a burden for the user who must overly rely on her memory to make sense of encountered information and get a sense of context out of it, which creates few opportunities for insight and contributes to unnecessary cognitive load.

I describe this problem as pertaining to continuity in the information space, among which the most salient challenge areas are:

1. Navigation through multiple data sets with incompatible structures.

2. Extracting entities beyond dedicated search tasks, and from any activity pertaining to the exploration, e.g., conversation, reading, writing.

Usability principles related to explorability have been proposed by Dörk in terms of *orientation*, *visual momentum* and *serendipity* [Dörk et al., 2011]. Both sets of principles differ in that Dörk's is formulated as applied features of exploration support, while mine attempts to define a theoretical representation of an explorable information space. The principle of *orientation* is shared in both proposals. *Visual momentum* matches closely my proposed principle of *direction*, but focuses on interface features, such as *animated transitions*, *zoomable interfaces* or *detail-on-demand*, while I describe the principle of *Direction* more generally. Finally, when considering properties of an explorable information space, I believe opportunities for *serendipity* should be an intrinsic result of being able to orient and direct oneself to reveal outlying or unusual results, defeating the need for its implementation as an added feature. For that reason, I propose instead the principle of *continuity*, which allows to further generalize the goal of an explorable information space.

These three properties of explorability do not exhaustively address all aspects of the main research problem but provide a comprehensive exploration of the intended design space as they demarcate the research so that it is possible to conduct within the scope of a doctoral thesis.

The research presented in this dissertation is the result of a highly multidisciplinary collaboration, spanning the fields of Human-Computer Interaction, Information Retrieval, and Interaction Design. The focus of this research can be described as Interaction Design applied to Entity Search. This dissertation does not address Semantic Web challenges or problems pertaining to the organization of information. It anticipates the availability of entity-oriented information, ideally in the form of an independent index of the web [Lewandowski, 2014], and utilizes Information Retrieval methods in the development of prototypes. Such methods enable the use of large but closed sets of indexed data, allowing the study of devised interaction techniques.

1.3 Research Questions

In the previous section, we have introduced three properties of explorability as research areas through which we address entity affordances for explorability that are *Direction*, *Orientation* and *Continuity*. These three areas have been operationalized into three respective research questions around which the work presented in this dissertation is articulated.

RQ1 : *How can entity-based querying benefit information exploration?* Conventional search methods usually use the frequency of occurrence of a phrase in a text as a proxy for relevance. In such a paradigm, text-based queries are an obvious way of formulating a search intent. In a context of entity-oriented information, interacting with entities as the primary method to express a user's search intent seems to have many benefits. Among those, the possibility to readily use found information as part of a query, therefore supporting query formulation through reliance on recognition over recall [Hearst, 2006]. However, expressing a user's intent through entity interaction may lack the precision of a well-formulated text-based query required to define a desired end-result. Thankfully, exploration implying non-clearly defined goals and an unfamiliar context, we can expect entity interaction to enable effective methods for providing a direction to the search. Therefore, RQ1 leads to the design of such methods, and their evaluation to determine whether entity-based interaction yields effective methods for expressing search directions, and can benefit information exploration over conventional text-based querying methods.

RQ2 : *How to demarcate and visualize a coherent information space through entity-based affordances?* Enabling orientation in the information space requires a map, which in itself is a considerable challenge. Common understanding of information as a space usually substitutes spatial distance for some conceptual distance, e.g., semantic proximity or adjacency. However, there are usually too many potentially useful components for computing such conceptual distances and to allow for a usable map-like projection. For example, two pieces of information could be considered as close from the viewpoint of their general topic, but extremely different in their approach, and they might have been published in different centuries but share a common geographic origin. There is virtually no limit to potential classifiers that would allow for some absolute mapping of information.

Orientation in the information space requires to work around such a limitation. In accordance with our design goals, we wanted to investigate user-driven methods to demarcate information of interest. More specifically, we were interested in investigating methods derived from entity-based querying to define mapping criteria and demarcate a corresponding information space.

RQ3 : *How to benefit from entity-based interactions for exploration beyond self-contained systems?* We have seen that entity-oriented information structures already serve the purpose of linking together heterogeneous information sets, thus partially addressing the continuity issue. I was more interested in addressing continuity from the viewpoint of explorers, whose activity is potentially long-term and might not be contained within one single convenient digital application. Search activities might be conducted on a search engine, but insights might come from re-reading an old email, asking a colleague for her opinion, or watching a seemingly unrelated video. We wanted to address continuity by exploring possibilities for extending any benefits of interactive entities beyond the confines of a dedicated application, which requires a proactive approach that relies on monitoring multiple aspects of a task in progress, and finding contextual information for immediate or later use.

1.4 Research Contributions

	RQ1	RQ2	RQ3
Publication I	•		
Publication II	•		
Publication III		•	
Publication IV		•	
Publication V			•
Publication VI			•
Publication VII			•
Publication VIII	•	•	•

Table 1.1. Distribution of the eight publications compiled in this dissertation with respect to the three research questions.

The research presented in this dissertation consists of an extensive design exploration of the topic at hand, and provides a variety of novel interaction techniques that have shown to support information exploration, and together demarcate a paradigm for future information practices. The research process has yielded eight publications, including seven design cases, all validated through user experiments, and a position paper. As shown in table 1.1, together they provide answers to all three research questions.

Publication I – Designing for Exploratory Search on Touch Devices

We present an entity-based technique for directing exploratory search on touch devices. We introduce ExplorationWall, a prototype exploratory search system that implements such a querying technique and present an experimental user study comparing the effects on search performance and behavior of such technique over a baseline system replicating conventional search engines when using portable touch devices. [Klouche et al., 2015].

Publication II – IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search

We introduce IntentStreams, a prototype that implements a technique based on user-intent modeling for directing exploratory search, and parallel search streams which enable visualization of simultaneous search sessions. We then present an experimental user study comparing the effects on the qualities of the search trail yielded by such technique over a baseline system replicating conventional search engines [Andolina et al., 2015b].

Publication III – Visual Re-Ranking for Multi-Aspect Information Retrieval

We present a visual querying technique based on multiple entities that represent result relevance and density on a map, and a technique to navigate the map by pointing at it, which triggers according re-ranking of the results. We then present an experimental user study comparing the effects on perception and retrieval over a baseline system replicating conventional search engines [Klouche et al., 2017].

Publication IV – QueryTogether: Enabling Entity-Centric Exploration in Multi-Device Collaborative Search

We present a prototype system designed to support co-located multi-device collaborative exploratory search through finding and sharing entities. We then present an experimental user study comparing the effects on participation, work-distribution and finding common ground over a baseline system replicating conventional search engines used collaboratively [Andolina et al., 2018a].

Publication V – InspirationWall: Supporting Idea Generation Through Automatic Information Exploration

We introduce InspirationWall, a display that leverages speech recognition to enhance ongoing idea generation sessions with automatically retrieved entities that relate to the conversation. We then present an experimental user study showing the effects of such apparatus on idea generation over time [Andolina et al., 2015a].

Publication VI – Investigating Proactive Search Support in Conversations

We study how a spoken conversation can be supported by a proactive search agent that listens to the conversation, detects mentioned entities, and proactively retrieves and presents related information. We then present an experimental user study showing how such proactive search agent augment conversations and affect topical structures [Andolina et al., 2018b].

Publication VII – Proactive Recommendation in Context: From Relevant Items to Actionable Entities

We present the design and implementation of an entity-centric proactive system that makes entity recommendations by capturing users' digital context. We then investigate whether the approach can effectively support everyday digital tasks by providing recommendations that have a concrete influence on users' tasks [Andolina et al., 2019].

Publication VIII – From Hyperlinks to Hypercues: Entity-Based Affordances for Fluid Information Exploration

We introduce the concept of Hypercue, a complement to hyperlinks in the form of an interactive representation of real-world entities (e.g., persons, places, concepts) providing personalized access points to information. The main contribution is a design template describing the Hypercue, which consists of a minimal set of affordances that ensure all important features for supporting exploratory search can be addressed [Klouche et al., 2018].

1.5 Structure of the Dissertation

The dissertation is articulated as follows:

Chapter 2 provides useful background information for the reader, as well as an overview of the closest work related to my research and of the notions that together demarcate and motivate the research.

Chapter 3 describes the methods applied for this research as components of *Research Through Design* methodology and *Experimental Research*, and provides details on ethical considerations.

Chapter 4 presents the research regarding entity affordances for directing exploration and addresses RQ1: *How can entity-based querying benefit information exploration?*. The chapter summarizes the systems developed in Publications I and II, overviews their respective evaluations, then highlights their joint contributions through their main findings.

Chapter 5 presents the research regarding entity affordances for orientation and addresses RQ2: *How to demarcate and visualize a coherent information space through entity-based affordances?*. The chapter summarizes the systems developed in Publications III and IV, overviews their respective evaluations, then highlights their joint contributions through their main findings.

Chapter 6 presents the research regarding proactive entity recommendation for continuity of the information space and addresses RQ3: *How to benefit from entity-based interactions for exploration beyond self-contained systems?*. The chapter summarizes the systems developed in Publications V, VI and VII, overviews their respective evaluations, then highlights their joint contributions through their main findings.

Finally, chapter 7 reflects on the findings of the research and presents as the main contribution a design template first presented in Publication VIII, a minimal set of affordances that ensure all important features for supporting exploratory search can be addressed. Additionally, limitations of the research and directions for future work are discussed.

2. Background

The research presented in this dissertation consists primarily of an extensive design exploration. This chapter provides an overview of the notions that together demarcate and motivate the design space of our research and describes these notions through the most closely related work.

2.1 Positive Information Practices

A large body of work builds upon what seems to be a human propensity to consider information as a space in which we move, progress and discover. Bates' berry-picking approach to search [Bates, 1989] acknowledges and describes the evolution of the cognitive model of a person as she goes through the search process, which happens bit by bit instead of in a linear fashion describe users' and describes such information-seeking behavior through the metaphor of a physical journey. The information foraging theory [Pirolli and Card, 1999] posits that human natural information-seeking behaviors use the same evolutionary mechanisms formerly used to find food.

In the "information flaneur", Dörk and colleagues go deeper into the physical metaphor and compare information spaces to the 19th-century city in terms of growth, cultural significance, and being the place for social struggle and negotiation. Information behaviors are not solely motivated by what they call negative approaches, i.e., information needs in the form of a knowledge gap to be filled, but include creative, participatory and serendipitous motivations, referred to as positive approaches to information practices [Dörk et al., 2011]. Such information-seeking behaviors are investigated and exemplified in the work of Thudt and colleagues [Thudt et al., 2015] through the study of a variety of search patterns of library patrons. Analysis of these patterns leads them to consider book search, in turn, as a creative process, through its social aspects, and as a serendipitous exploration experience.

Thudt and colleagues address the importance of serendipity in support of finding relevant information in exploratory scenarios [Thudt et al., 2012]. With *The Bohemian Bookshelf*, they explore support for serendipity through information visualization by taking a cue from the physical action of "browsing the shelves". As a result, they have devised five design goals to promote serendipity through information visualization: (1) *multiple visual access points*, *highlighting adjacencies*, *flexible visual pathways*, *enticing curiosity* and *playful exploration*. These five principles have then been implemented into a search system, *The Bohemian Bookshelf*, that was embraced by library visitors. Such an example encourages the conceptual transposition of properties of physical activities into their digital counterparts. The generalization of such an approach is at the center of the research work presented in this thesis.

Implications of such work for research offer design goals for fostering or enabling these experiences by considering various explorability principles, e.g., orientation, visual momentum, and opportunities for serendipity, and bridging gaps between information spaces, contexts, and conceptual levels by exploiting scalable or generalizable rules and common patterns. Such models have all contributed to shift the emphasis from a mostly technical consideration of information retrieval toward information practices as human processes [Kerne and Smith, 2004].

2.2 Direct Manipulation and Fluid Interactions

Direct manipulation describes the mode of operation of reactive interfaces that continuously represent the objects and actions of interest and rely on physical action instead of complex syntax [Shneiderman, 1997], an interaction paradigm in which digital representations of objects behave as objects themselves [Shneiderman, 1993]. Direct interaction with these objects is enabled by reducing indirections between input and output spaces. For example, the touch-sensitive layer of a touch device is confounded with its display and calibrated so that inputs are registered precisely at the display location. The paradigm relies on immediate visible effects allowing rapid course adaptation [Hutchins et al., 1985]. The move from devices using the command line towards mouse input and touch-based interaction are two important steps in this direction, allowing more closely the human action to happen where the effect takes place in the machine. Direct manipulation is known to improve user satisfaction through multiple effects: Facilitated learning for novices, improved efficiency for experts, overall more confidence and less anxiety through continuous feedback and better predictability [Shneiderman, 1997].

In advocating for direct manipulation, Shneiderman pits designers who are proponent of autonomous, adaptive, intelligent systems against advocates of user-control, responsibility, and accomplishment [Shneiderman, 1995]. By emphasizing control over a system, users become responsible for their actions. Applied to information systems, we immediately see how conventional search engines, which are suspiciously reminiscent of the command line, deny such responsibility to users.

When we discuss interactions and user interface design, fluidity is often cited as a goal [Guimbretière et al., 2001, Ramos and Balakrishnan, 2003]. White and Roth [White and Roth, 2009] mention fluid interactions as an important feature of future search systems when discussing novel interaction paradigms. They link that notion to human-machine symbiosis and interactions through fluid hand gestures, citing the fantasy user interface used by anticipatory investigators in the movie *Minority Report* as an example of what a truly fluid interface could look like. However, that notion is not theoretically defined and is generally used while relying on the reader's intuitive understanding of the metaphor, something that flows continuously, naturally making its way around obstacles and adapting its pace to the environment.

Elmqvist and colleagues propose a satisfying operational definition [Elmqvist et al., 2011], avoiding the difficulty of defining fluidity theoretically by focusing on the properties we can expect from fluid systems. These properties are grouped into three sets:

1. Fluid interactions support direct manipulation.

2. Fluid interactions promote flow[1].

[1]Flow is a mental state induced by immersion in one's activity, characterized by a loss of sense of time. The main actionable property for inducing flow relies on letting users feel in control, and employ just the right amount of skills to let them progress in their tasks at a pace that will feel neither too slow nor too fast, accommodating a person's continued and deepening enjoyment as skills grow [Nakamura and Csikszentmihalyi, 2014].

3. Fluid interactions minimize the gulfs of action[2].

A limitation of this definition is that these properties are largely overlapping, and they seem contained in Shneiderman's understanding of direct manipulation, and advocacy for comprehensible, predictable and controllable user interfaces; however, it provides various approaches to the implementation of such goal.

2.3 Entity Search

In web searches, a majority of emitted queries pivot around a specific entity [Jansen and Spink, 2006, Pound et al., 2010, Garigliotti and Balog, 2018]. In common search engines, queries pointing towards an entity, e.g., "Who is Tom Hanks?" or more simply "Tom Hanks", will usually trigger a first result that points to the information source with the most general information about the entity, typically the corresponding Wikipedia entry. As the web still lacks a definitive repository of entities, Wikipedia is often used as such a repository, as it provides information on most entities.

Google Search provides for most entity-based queries not only a relevant entry but what they call "the knowledge graph", an infobox showing relevant information about the central entity in the query, with recommended related entities. For example, in the case of an actor: name, age, and lists of movies and co-stars. Miliaraki et al. [Miliaraki et al., 2015] studied the behavior of users of Yahoo Spark, a system that recommends related entities alongside Yahoo Search results; the users take advantage of the system to engage in explorative entity search by discovering information through successive clicks on recommended entities. Such cases exemplify why entity-search is considered an ideal paradigm for exploratory search and an important topic in information retrieval and semantic web communities. A large body of recent research work addresses challenges regarding the computation necessary for entity search, such as the finding and ranking of related entities, matching entities with occurrences in free text queries and completion of entity lists based on given entity examples. However, as techniques improve, it is difficult to find research addressing interaction techniques that enable end-users to access and benefit from such rich informa-

[2] The gulfs of action are a notion introduced by Donald Norman [Norman, 2013], who uses it to describe the gap between a user's expectation of a system and the system's actual state.

tion in a wide variety of activities, as examples usually target very specific scenarios and tasks. Several such entity-based exploration systems have been designed to support expert investigators in making sense of a corpus of documents [Bier et al., 2008, Stasko et al., 2008, Carmel et al., 2012].

The goal of entity-based fluid information exploration requires substantial thinking about the way we display and interact with entities and come up with fundamental principles that are generalizable to any search contexts (e.g., academic publications, social media, movie database and personal emails).

2.4 Features of Exploratory Search Systems

The work of White and Roth on Exploratory Search provided a precious frame to our understanding of how to support exploration [White and Roth, 2009]. Their work includes a list of features of exploratory search systems that exemplify essential aspects of supporting information exploration and provide a useful overview of the various directions in which the work presented in this dissertation has sought to address the problem at hand.

Support for Querying and Rapid Query Refinement
Search tasks are commonly addressed by inputting queries in a search system, which then yields a set of related results. However, conventional text-based queries are mostly user-defined. Relying on the user's existing knowledge to formulate satisfying search directions limits the range of incrementation in the iterative exploration process [Teevan et al., 2005]. Support for querying is commonly addressed by providing the user with ideas for new queries or additional terms.

Facets and Metadata-Based Result Filtering
Being able to navigate a large result set according to personal needs and preferences is a central requirement of fluid information exploration. That is why this ability – to narrow down such results according to a variety of criteria that are representative of what is available in the data and complementary enough to provide a meaningful choice of search directions – is an important feature to support.

Facets and metadata-based parameters are an attempt to structure information by linking documents semantically through common features (e.g., an author, a title, a date or a location). Faceted browsing has shown to be effective in supporting sense-making. Various visualization techniques have been explored through a variety of work [Lee et al., 2009, Smith et al., 2006], supporting the exploration of a collection of documents through meta-data-based commonalities.

Similarly, entity search is the ideal paradigm for result filtering, given the richness and complexity of readily-linked data. From an initial query entity, a system would retrieve the most central neighboring concepts or elements and provide them as related entities to choose from. The initial result set can then be narrowed down or re-ranked with respect to the relatedness or dependency of each element to the chosen related entity.

Leverage of search context

A substantial part of context can already be harnessed by accessing contextual data provided by sensors (e.g., GPS signal or personal account information). From our interaction design perspective, and in light of our goal of increasing user control, I am more interested in techniques enabling inference of context through users' input, either explicit or implicit. Inferring context from explicit user input includes detecting relevant relations from the way she organizes stored information. Useful implicit inputs can include what the user has been reading, oral conversation, or physiological signals [Barral et al., 2018], which can be used by the system to infer suitable relations or topics.

Visualizations to Support Insight and Decision Making

Interactive information visualization is an important tool for sense-making. Being able to encode data visually and to play with various parameters is a powerful way of discovering trends, understanding relationships, gaining insight from the data and ultimately informing decisions. Entities in knowledge graphs generally make for inspiring material regarding visualization techniques such as node-and-link diagrams and adjacency matrices.

Support for Learning and Understanding

As it is necessary to offer some result-filtering ability for the user to take better advantage of a large set of results by narrowing down a list, it is also important to provide the user with access to more general knowledge when necessary. Support for learning and understanding implies that a user is given the means to find information that is adapted to his current level of understanding. This is typically achieved through the recommendation of related material. For example, any modern browser or eBook reader provides the ability to look up the definition of a word or to link a concept with its corresponding Wikipedia entry.

Support for Collaboration
Collaborative information exploration is a common strategy to tackle large information spaces through the sharing of ideas and allocation of search tasks [Hearst, 2014]. Collaboration can take multiple forms, with settings in which collaborators either share or do not share the same space (i.e. co-located or distributed collaboration), either synchronously or asynchronously.

Histories, Workspaces, and Progress Updates
Information exploration is a sense-making activity [Marchionini, 2006]. As such, it is open-ended, potentially long-term, and changes continuously as the information needs evolve [Pirolli and Card, 2005]. The process often produces long and complex search trails with multiple branches and revisits. In this context, a user needs to be able to take advantage of previously encountered information and to keep track of past activity to more efficiently recognize new and interesting information.

Support for Task Management
As information exploration is potentially long-term, users need to have the ability to interrupt and resume their activity and to carry it over time and across devices. This requires the ability to not only save selected information but to provide future access to whole workspaces, including histories and information configuration with which a user has engaged.

2.5 Visualization of the Information Space

The following section provides an overview of closely related work involving data visualization. First, with regards to directing the exploration, the section covers various techniques for visual information retrieval and multi-aspect search, i.e., search involving multiple simultaneous criteria. Then, pertaining to orientation in the information space, I overview techniques for visualizing large document collections and examples of user-driven visualizations.

2.5.1 Visual Information Seeking

Information spaces can be huge and thus hard to comprehend. However, visualizing the space and allowing the user to directly interact with and manipulate objects in the space facilitates comprehension. For instance, when the results of actions are shown immediately and when typing is replaced with pointing or selecting, exploration and retention increase while errors decrease [Shneiderman et al., 2009]. For information seeking, the following visualization and interaction features are of particular importance [Shneiderman, 1994]: (a) dynamic querying for rapid browsing and filtering to view how results change; (b) a starfield display for the immediate, continuous, scalable display of result sets as different queries are processed; (c) tight coupling of queries to easily use the output of one query as input to another [Ahlberg and Shneiderman, 1994]. For instance, a user study indicates that dynamic querying significantly improves user response time and enthusiasm. Using such techniques, systems like FilmFinder [Ahlberg and Shneiderman, 1994] support querying over multiple varying attributes such as time, while showing the changing query results in the context of the overall data. Another example is found in VisGets [Dörk et al., 2008], which augments search and exploration on the Web with a variety of coordinated visualizations, providing not only a multidimensional overview of the information space, but also the visual means to precise the query and filter the data.

User studies also indicate that user interfaces that show the result list together with an overview of the result categories encourage a deeper and more extensive exploration of the information space [Kules and Shneiderman, 2008], especially when the system allows relevance feedback to be given on such categories to direct the exploration [Ruotsalo et al., 2013b, Ruotsalo et al., 2015].

2.5.2 Multi-Aspect Search

In multi-aspect search the information need of the user consists of more than one aspect or query simultaneously. As a consequence, an item in a collection needs to be ranked differently based on its multiple attributes. The Graphics, Ranking, and Interaction for Discovery (GRID) principles and the corresponding rank-by-feature framework state that interactive exploration of multi-dimensional data can be facilitated by first analyzing one- and two-dimensional distributions and then by exploring relationships between the dimensions, using multi-dimensional rankings to set hypotheses and statistics to confirm them [Seo and Shneiderman, 2005]. However, comparing, analyzing and relating different ranks is difficult and requires an interactive visualization that supports the various requirements identified by Gratz et al. [Gratzl et al., 2013]. For example, Polaris [Stolte et al., 2002] is a visualization tool for exploring large multidimensional databases. The system allows the user to drag relational aspects from the database schema onto the display area and select context-relevant display specifications to generate a variety of rich visualizations.

Multi-aspect search support is provided in Song et al. [Song et al., 2012], with the proposal of a strategy for a multi-aspect oriented query summarization task. The approach is based on a composite query strategy, where a set of component queries are used as data sources for the original query. Similarly, Kang et al. [Kang et al., 2012] propose a multi-aspect relevance formulation, but in the context of vertical search.

LineUp [Gratzl et al., 2013] is an interactive visualization that uses bar charts to support the ranking of objects with respect to multiple heterogeneous attributes. Stepping Stones [Das-Neves et al., 2005] visualizes search results for a pair of queries, using a graph to show relationships between the two sets of results. Sparkler [Havre et al., 2001] allows us to visually compare results sets for different queries on the same topic. Tilebars [Hearst, 1995] visualizes the frequency of different words in various sections of documents as a heat map and ranks the documents accordingly. Similarly, HotMap uses a two-dimensional grid layout to augment a conventional list of search results with colors indicating how hot (relevant) specific search terms are with respect to the document [Hoeber and Yang, 2006b]. Ranking cube [Xin et al., 2006] is a novel rank-aware cube structure that is capable of simultaneously handling ranked queries and multi-dimensional selections. RankExplorer [Shi et al., 2012] uses stack graphs for time-series data. Techniques for incomplete and partial data have also been proposed [Kidwell et al., 2008]. TreeJuxtaposer [Munzner et al., 2003] was primarily devised to compare rankings.

For document collections, the vector space model could be used, such that each document and search query is a vector in a multi-dimensional space, each axis is a term, and the document position is determined by the frequencies of each term in that document (e.g., [Raghavan and Wong, 1986]). Visualizations of such a model could aid understanding of the document space, but more research is required, particularly for user-driven approaches that allow the user to specify the dimensions of interest [Olsen et al., 1993].

2.5.3 Visualization of a Document Collection

Various visualizations have been proposed for large document collections [Kucher and Kerren, 2015]. Most of these techniques adopt the visual information seeking mantra [Shneiderman, 1996] to provide an overview at first and details only on demand. The documents are often visualized on a 2D plane, in the form of a map based on a similarity metric. Higher-level entities, such as topics, are also displayed on the map for immediate and better understanding of the document space organization.

Document Atlas [Fortuna et al., 2005] uses Latent Semantic Indexing and multi-dimensional scaling (MDS) to extract semantic concepts from the text and position the documents with respect to the concepts. Document densities around concepts are visualized as a heat map. On mouse hover, common keywords in the area are listed, and on zoom in, more details are shown.

Self-Organizing Maps have also been used by systems like WEBSOM [Kaski et al., 1998] and Lin's maps [Lin, 1997] to position the documents on the 2D plane. WEBSOM also suggests areas in the map that could be relevant to the user's search query. Lin's maps are further split up into regions whose area indicates the number of documents with specific related terms.

Other techniques visualize the documents as glyphs to indicate additional inter-document relationships and metadata on the map (e.g., [Rohrer et al., 1998, Miller et al., 1998]). Various metaphors have also been adopted; examples include the terrain metaphor, in which dense regions in the map are seen as mountains with valleys in between [Boyack et al., 2002, Wise et al., 1995]; the galaxy metaphor, in which documents are seen as stars in different constellations (document clusters) [Hetzler and Turner, 2004]; and the physical metaphor, in which documents are considered to be moving particles and the inter-particle forces move similar documents closer to each other and dissimilar documents apart [Chalmers and Chitson, 1992]. Visualizations with two dimensions and meaningful axes (e.g., categories vs. hierarchies [Shneiderman et al., 2000], query results vs. query index [Ruotsalo et al., 2016], production vs. popularity [Ahlberg and Shneiderman, 1994]) have also been proposed.

ResultMaps [Clarkson et al., 2009] takes advantage of pre-existing ontologies used in digital libraries to create a hierarchical view in the form of a tree-map [Johnson and Shneiderman, 1991]. Such a technique allows for consistent representation of a given document collection. Integrated within the result page of a digital library, found or explored documents can be highlighted on the visualization, providing the user with a growing sense of familiarity over areas of interest.

These visualizations provide an overview of the entire document collection, but they do not allow the user to direct and focus the exploration as required. A user-driven rather than a data-driven technique could be more helpful when searching for documents relevant to multiple keywords. To that end, such a technique should visualize the ranking of documents with respect to multiple keywords so the user can easily judge the relevance of documents to each of the keywords of interest [Olsen et al., 1993]. However, most of the current techniques only visualize whether a document is relevant or not to a keyword using set visualizations [Alsallakh et al., 2014], without showing the document's degree of relevance to each keyword.

2.5.4 User-driven Visualization

VIBE [Olsen et al., 1993] is one of the most well-known user-driven multidimensional ranking visualization for large document collections. To indicate the subspace of interest, the user first enters two or more query terms, known as "points of interest" (POIs). POIs are then shown (as circles) on a 2D plane, together with documents (as rectangles) related to at least one POI, forming a map. The position of each rectangle indicates the relevance of the corresponding document to each of the POIs. The size of a rectangle indicates the relevance of that document to the search query. Citation details of documents selected from the map are listed; clicking on an item in the list opens the full document. Any time a POI is added, removed or moved, the map is updated accordingly. However, regions of the map with numerous close-by documents are not easily detectable because the rectangles are not color-filled; using semi-transparent color-filled shapes reduces overplotting [Matejka et al., 2015] and facilitates the perceptual ordering of different regions in the map by their density [Mackinlay, 1986]. Also, documents are not re-ranked as the user navigates over the map.

Variants of VIBE include: WebVIBE [Morse and Lewis, 1997], in which POIs act like magnets that attract documents containing related terms; VR-VIBE [Benford et al., 1995], which visualizes the space in 3D (for more space to view documents between POIs) and depicts relevance by color; and Adaptive VIBE [Ahn and Brusilovsky, 2009], in which POIs are query terms (as in VIBE) but also user profile terms that are automatically extracted from user notes.

Similar to VIBE, GUIDO [Nuchprayoon and Korfhage, 1994], DARE [Zhang and Korfhage, 1999] and TOFIR [Zhang, 2001] also allow users to specify POIs and display documents based on their relevance to the POIs. However, in GUIDO each POI is an axis (not an icon on a 2D plane) and documents are positioned based on their absolute rather than relative distances from the POIs. In DARE and TOFIR, relevance to POIs is indicated by both distance and angle.

Other user-driven systems, like combinFormation [Kerne et al., 2006], TopicShop [Amento et al., 1999] and InfoCrystal [Spoerri, 1993], retrieve and display search results related to user-defined keywords but do not visualize the results' multi-dimensional ranks. Similarly, HotMap [Hoeber and Yang, 2006b] supports a weighted re-ranking of the search results, but without leveraging a graphical interactive approach for specifying the weights. WordBars [Hoeber and Yang, 2006a] also supports re-ranking of the search results, but uses additional terms extracted from the search results rather than relying on the query terms.

This section aimed to exemplify the extensive previous work surrounding information visualization and the various purposes to which it can be applied in the context of this research. It would not be complete without mentioning the work of Hinrichs and Forlini defending *sandcastles* [Hinrichs and Forlini, 2017], in which they make the case for information visualization as an exploratory process in itself that can potentially yield insights through its inherent interdisciplinarity and aesthetic provocations, independently from its consideration as a means to an end.

2.6 Digital Activity Monitoring

Continuity between various information spaces, for example, harnessing insights from a read or a conversation for immediate or later use, requires proactive approaches involving monitoring of a user's digital activity. Research on digital activity monitoring and prediction of user behavior has typically focused on large-scale tracking, e.g., based on what people are sharing on social media [Zhu et al., 2013, Yang et al., 2015]. This mass monitoring approach has some important drawbacks, including loss of privacy and lack of trust for the system [Chaudhry et al., 2015]. However, some recent work has studied technologies for individual monitoring of personal data, putting the collection and analysis of the data into the hands of the individuals themselves [de Montjoye et al., 2014, Sjöberg et al., 2017].

Most of the approaches mentioned so far have focused on monitoring specific applications or other limited data sources. However, recent work [Vuong et al., 2017a, Vuong et al., 2017b] has explored using screen monitoring, which captures the entire visual content of the computer screen for task recognition. Latent Semantic Analysis [Deerwester et al., 1990] with a simple bag-of-words data representation was found to be the most effective to detect users' tasks and helpful for proactive information retrieval. This tracking "inside the screen" paradigm has the benefit of being more general, as any visually communicated information can potentially be captured, and utilized for building a richer task model. An approach similar in spirit, but more limited, is described in [Gyllstrom and Soules, 2008] where seen text snippets are associated with files opened at the same time.

2.6.1 Using Background Speech for Interaction

Speech-based interaction has been thoroughly studied in the literature. However, the interest in speech-based systems seems to have risen again in recent years, probably due to the recent advances in automatic speech recognition [Negri et al., 2014]. In particular, a large body of work focuses on a dialogic mode of interaction [McTear, 2002] where users communicate with the system using natural language. Commercially available examples include *Apple's Siri*, *Microsoft's Cortana*, and *Google Now*.

Less investigated is the use of background speech for interaction. One example is *Ambient Spotlight* [Kilgour et al., 2010], which uses speech recognition during meetings to search for desktop documents and puts them in a folder associated with the calendar entry related to that meeting. Other systems use background speech to retrieve words and other kinds of visual stimuli to support a creative conversation [Shi et al., 2017]. As opposed to those systems, which are designed to support creative conversations where even misrecognition and random results may lead to useful stimuli [Kirsh, 2014], we investigate how to support more generic conversations by proactively retrieving richer sources of information, such as documents, from the Web.

An important study related to our work is that of McGregor and Tang [McGregor and Tang, 2017]. The aim of their study was to understand how well a speech-based agent could detect useful actions during formal meetings. Although the study used a simulated system to create a best-case scenario, results showed that extracted action items failed to fit with the meeting or gave an incorrect summary of what was being discussed or what the participants intended. A different approach was that of McMillan et al. [McMillan et al., 2015]. Their study suggested that a continuous speech stream, rather than containing directly actionable items, can be used to identify users' next actions such as searches. This result inspired our research, as it means that regardless of the limitation of current automatic speech recognition technology, many useful words that would likely be used for a search could still be recognized. In this study, we aim to understand whether performing those searches proactively during conversations could effectively enrich those conversations.

2.7 Summary

The overview of *positive information practices* presented in this chapter has shown how activities pertaining to information seeking are not to be limited to the bridging of a gap or the fulfillment of a need but can be led by curiosity and serendipity. The design goal of explorability through orientation and continuity described by Dörk and colleagues [Dörk et al., 2011] inspire us a mental picture of a wanderer engaging with an information space as she would with a city, orienting herself and becoming familiar enough with her surroundings that she starts noticing unadvertised changes. In today's reality, conventional search tools will only help her find something for which she already has an interest in and some knowledge of, and recommender tools remove her from the active role required in exploration. Is it then possible to think of tools that would enable or facilitate such behavior?

As this chapter has shown, fluidity in interactions results from the ability of a user to feel in control of a process despite the potentially high level of abstraction of such process. When looking for examples of interactions qualified as fluid, we realize all examples focus on replicating properties of physical interaction to take advantage of a user's existing experience of the world as a means to infer the correct sequence of action to take towards the desired goal. Such observation encourages us to identify properties of the physical space that can be emulated in the information space to enable exploratory behaviors that do not benefit from conventional search and recommendation approaches.

Then the chapter covers the use of entities for search and exploration. Entities create a double opportunity: On the one hand, they enable the organization of information into a continuous metric space. On the other hand, they offer a flexible unit of information with which to interact in novel ways. These two aspects seem to simultaneously provide the field and the vehicle for fluid exploration, which encouraged us to make entity interactions the technical focus of our research. The features of exploratory search systems summarized in this chapter offer precious guidelines when designing for information exploration. However, we want to avoid the trap of implementing these features individually into bloated interfaces that would largely defeat our goal of fluid interactions. Instead, we will be using these features as a lens to assess the validity of our designs and their potential for exploration.

Subsequently, the chapter showcases a variety of significant works pertaining to information visualization. Through examples of how a document collection and an information space of interest can be demarcated and explored visually.

Finally, the chapter presents research work pertaining to digital activity monitoring as it is used as a means to transpose the property of continuity from the physical space to the information-seeking practices.

Together, these sections provide the reader with the necessary background to understand the variety of techniques used in the present research as well as the preliminary information required to put the present research in its proper context.

3. Methodological Considerations

This chapter overviews the research strategies and methods used, as well as the ethical concerns and how they were addressed.

3.1 Research Through Design

The work presented in this dissertation is an exploration of a design space through seven projects reported in publications I to VII, each of which follows the structure of *Research Through Design* as established by Alain Findeli and colleagues [Findeli et al., 2008], as they all feature clear *Research For Design*, *Research About Design* and *Multidisciplinary* components.

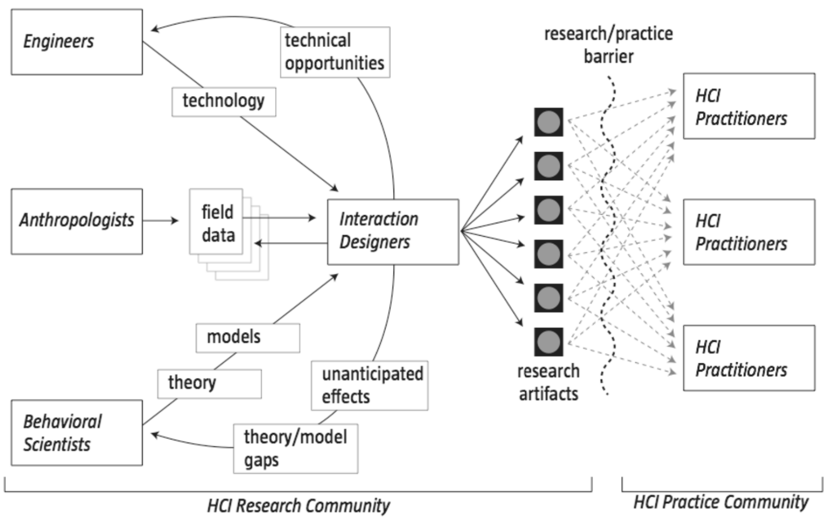


Figure 3.1. Model of the role of the interaction design researcher among other HCI researchers, emphasizing the production of research artifacts as units of analysis.

Applied to Human-Computer Interaction, the role of the designer engaged in such methodology is accurately described by Zimmerman: Interaction design researchers integrate the models and theories from behavioral scientists, with the technical opportunities demonstrated by engineers to produce artifacts. These artifacts enable speculative exploration influencing both the research and practice community [Zimmerman et al., 2007]. The model is illustrated on Figure 3.1.

Research For Design

Each project starts with a thorough literature review around the challenge at hand, aimed at identifying a clearly demarcated gap in knowledge, and inform the design of an information system and its user interface. The specifications for the system are in the case of this research put in the form of a scenario exemplifying a detailed use case. The scenario is used as a designed tool at various stages of the design process as it provides multiple benefits exemplified by Caroll [Carrol, 1999]. In the ideation phase, it helps define the design intent through typical use cases to be addressed. In a multidisciplinary context such as the one from which the present research stems, a scenario is a precious tool for involving collaborators with different fields of expertise, e.g., interaction designer and engineers in retrieval systems, by communicating an intent grounded in real-world use, then letting them infer resulting technical requirements and raise questions, potential difficulties or provide insights, therefore finding common-ground and ensuring converging work effort throughout the design, implementation and evaluation processes. This stage qualifies as *Research For Design* as it consists of drawing on available knowledge to produce an artifact.

Research About Design

Each of these artifacts then becomes itself a unit of analysis. It is evaluated following rigorous experimental research methods with the objective of producing new knowledge that is relevant to the design community, usually in the form of design principles that have been identified as providing benefits in the context of the challenge at hand.

Multidisciplinarity

All the research work reported in this dissertation is the product of highly multidisciplinary collaborations at the intersection of *Information Retrieval*, *Human-Computer Interaction* and *User Interface and Experience Design*. As evidenced by the generally high number of authors on Publications I to VII, corresponding projects are the outcome of the various domains of expertise being involved at every level of each project.

3.2 Experimental Research Methods

Findeli et al. emphasize the necessity for *Research Through Design* projects to be rigorous and stand up to usual scientific standards [Findeli et al., 2008]. We used experimental research methods allowing us to design tasks involving various conditions whose comparison would enable investigation of a set of hypotheses. To design a task or select a method most suitable to validate a hypothesis, we strived to maximize three features of measurement devised by McGarth: generalizability of the results across the population of users, precision of the results through control of confounding factors, and the realism of the context in which the experiment is conducted [McGrath, 1995].

To answer the research questions, we have opted for *controlled laboratory experiments* as a method for maximizing *precision* and control over confounding factors, thus enabling comparison of the experimental condition with a corresponding baseline condition, which would have been extremely difficult to do in the field. As McGarth explains it, each of the three features tends to interfere with one another, and the chosen strategy represents challenges the realism of the context as it constrains the location and the task. We strived to compensate such trade-off by designing naturalistic tasks grounded in the literature. Generalizability was assured through a sufficient amount of participants and systematic computation of the significance of the results, i.e., the probability that a difference between the conditions is not coincidental.

We opted for systematic *within-subjects* experiment designs, meaning that we relied on the same pool of participants for testing both experimental conditions, as opposed to *between-subjects*, which would imply distinct pools of participants for testing each condition. Within-subjects experiment designs require less overall participants while minimizing random noise introduced by individual personalities, however, they require some care to avoid participants to transfer acquired knowledge from the first tested condition to the next. To that end, we carefully balanced among participants the order of the tested conditions, as well as the provided topics to be explored.

3.3 Data Collection

Data collection was done in several forms:

3.3.1 Experimental and baseline conditions

Experimental and baseline conditions were prepared to allow *logging of user actions*, e.g., queries sent, typed, results received, displayed, consulted and saved. Despite needing the baseline condition to accurately replicate a realistic situation, we could not simply use conventional existing tools, as they would have utilized a different backend, i.e., data collection and retrieval algorithm, thus introducing noise. Therefore, to be able to compare interaction-based effects between conditions, we needed for our baselines to create custom systems with a conventional interface that used the same backend as in our experimental condition.

3.3.2 Videotaping

Videotaping enabled transcription of dialog and logging of occurrences of screen gaze in Publications IV, V and VI, which are summarized respectively in Sections 5.2, 6.1 and 6.2 of this thesis. Dialog transcription was done by hand using specialized software for qualitative experiments (ATLAS.ti) that facilitated the assignment of transcribed dialog to specific timecodes in the video. Screen gaze monitoring was also done manually. In the study reported in Publications V and VI, the setup conveniently involved laptop computers whose embedded camera right above the display provided a point of view that made occurrences of screen gaze obvious. In the study reported in Publication IV, the large public display to which gazes were monitored was large enough and positioned in a specific angle that allowed us to pick up such occurrences from the point of view provided by the general camera.

3.3.3 Usability Testing

Standard questionnaires were used to inform us of usability factors.

The Standard Usability Scale (SUS), seen in Figure 3.2, is a questionnaire described as a "Quick and Dirty" usability scale that provides a global view of subjective assessments of usability [Brooke et al., 1996]. SUS broadly covers the notions of effectiveness, efficiency, and satisfaction provided by a given system in the general context of the purpose in which it is being tested. It consists of a simple, ten-item scale, each item consisting of a Likert scale, i.e., a statement to which the participant indicates a level of agreement on a (usually) five-point scale. SUS was used in the experiments reported in Publications I, II, IV and VI, respectively summarized in Sections 4.1, 4.2, 5.2 and 6.2.

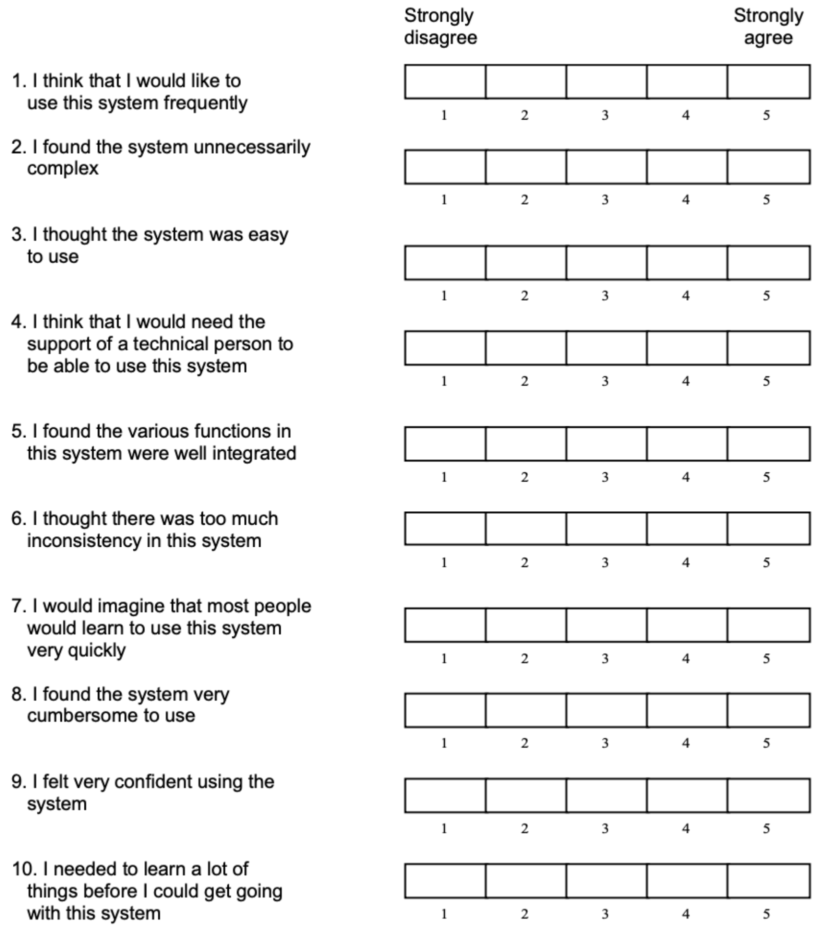


Figure 3.2. The ten items of the Standard Usability Scale (SUS).

The User Engagement Scale (UES), seen on Figure 3.3, is a multidimensional questionnaire, initially developed with thirty items (Likert scale) allowing to assess users' perceptions of six factors: the Perceived Usability (PUs), Aesthetics (AE), Novelty (NO), Felt Involvement (FI), Focused Attention (FA), and Endurability (EN) aspects of a system [O'Brien and Toms, 2010]. UES has proved useful in the contexts of exploratory search and interactive information retrieval

[O'brien and Toms, 2013]. In its revised short form [O'Brien et al., 2018], UES includes twelve questions addressing four factors: Focused Attention (FA), Perceived Usability (PU), Aesthetics (AE) and Reward (RW). UES was used in its short form in the experiments reported in Publications I, II, IV and VI, respectively summarized in Sections 4.1, 4.2, 5.2 and 6.2.

FA-S.1	I lost myself in this experience.
FA-S.2	The time I spent using Application X just slipped away.
FA-S.3	I was absorbed in this experience.
PU-S.1	I felt frustrated while using this Application X.
PU-S.2	I found this Application X confusing to use.
PU-S.3	Using this Application X was taxing.
AE-S.1	This Application X was attractive.
AE-S.2	This Application X was aesthetically appealing.
AE-S.3	This Application X appealed to my senses.
RW-S.1	Using Application X was worthwhile.
RW-S.2	My experience was rewarding.
RW-S.3	I felt interested in this experience.

Figure 3.3. The twelve items in the short form of the User Engagement Scale (UES) address four factors: Focused Attention (FA), Perceived Usability (PU), Aesthetics (AE) and Reward (RW).

ResQue (**Re**commender **s**ystems' **Q**uality of **u**ser **e**xperience) is a user-centric evaluation framework for recommender systems [Pu et al., 2011]. The prototypes developed in our exploration of entity-based querying techniques function as recommender systems, as they use a collection of items – in our case defined by the users as a means of inputting their intent and direct the search– to return a selection of related items. Such questionnaire was useful in assessing whether the returned elements matched users' expectations. SUS was used in the experiments reported in Publications I, II, IV and VI respectively summarized in Sections 4.1, 4.2, 5.2 and 6.2.

Usability was also assessed through semi-structured interviews performed after the session, to inform us on users' subjective experience and benefit from her potential insights. The *Semi-structured* qualifier refers to the fact we usually had a prepared bullet-point list of subjects to address, but we did not avoid digressions and informal conversation. Such interviews were used in the experiments reported in Publications I, III, IV and VI, respectively summarized in Sections 4.1, 5.1, 5.2 and 6.2.

3.4 Ethical Considerations

In all our experiments, participants were informed of our privacy guidelines upon joining. They were told that the data would be encrypted and stored on a secured server at the university, only to be used for research purposes. When applicable, participants were also informed of the use of Google and IBM services in the experiment. According to the term of services given by Google and IBM, the data would only be stored for a period of time sufficient to perform its analysis.

Prior to taking part in the study, participants had to sign a consent form stating the procedure of the study and data usage policy. Participants were informed that they were allowed to withdraw from the experiment at any time, in which case all resulting data would be removed from any respective server or storage device.

No experiment involved the collection of data without the participants' prior knowledge and consent. All research followed the ethical guidelines of the University.

4. Directing Exploration Through Entity Interaction

While document-based search relies heavily on frequency – the occurrence rate of a text-based query in the content of a document – entity-based data, or knowledge graph, is less content-focused, therefore the same technique does not apply. The theory of how to retrieve entities in a graph is well established and already common use in conventional search engines. These still rely on typed queries but use many techniques to identify entities within a query typed in natural language for disambiguation purposes and to improve search results.

Techniques involving direct interaction with entities also exist, however it mostly consists of following information paths provided by successive single recommended items, like following a path of hyperlinks. Such an approach can lead to enjoyable serendipity, but it lacks the means to yield the personalized results that a typed query can provide. As for typed queries, we already established that recalling search terms to formulate and refine queries, is not as easy as following or selecting visible items, especially when exploring unfamiliar information spaces.

We were interested in developing techniques for entity-based search that involve direct interaction with entities and are able to yield a scope of search results that is at least comparable to typed queries.

Entity-based queries consist of a set of one or several entities of interest that together express a user's search intent. As a result, such query yields a crop of new entities, ranked according to their overall relatedness to the query. Entities from the results, besides conveying potentially interesting information, can readily be added to the current query or be used as a new separate query. Entity-based querying offers support for rapid query refinement while enabling more personalized search directions than following single recommended entities.

The process of formulating queries through association of concepts is a compelling idea but to our knowledge, the literature does not provide observations on whether such approach is effective in enabling users to express and refine search intents, and can benefit an iterative process such as exploratory search, especially when compared to the conventional typed-queries approach. Therefore, validation of the use of entity-based querying for exploratory search first required testing of system implementing such features to understand how it affects performance and behavior when used in exploratory tasks.

This chapter reports on the design of two prototype systems that make use of different techniques for entity-based querying, showcasing the effects of these systems on search performance and user behavior in tasks pertaining to exploratory search.

The first one, summarized in Section 4.1 and reported in Publication I, allowed us to investigate the potential of entity selection for directing exploratory search. The second one, summarized in Section 4.2 and reported in Publication II, adds the possibility for relevance feedback and user modeling to explore the potential for user-driven recommendations in directing exploratory search. User-testing of both systems provided us with satisfying answers to RQ1 : *How can entity-based querying benefit information exploration?*.

4.1 ExplorationWall: Directing exploratory search through direct manipulation

Taking cues from the simplicity of the hyperlink, we looked for a technique that would allow someone to quickly react to any inspiring bit of information and be able to follow any encountered conceptual lead. As stated above, several search engines already suggest entities to follow ("you might be interested in...") but a succession of single entities does not rival with the scope of results that can yield more complex types of queries.

Therefore, we decided to investigate the results of querying by following multiple entities simultaneously. This technique allows for different kinds of concept associations. For example, using "Finland" as an initial query would yield a crop of various results, potentially including "Independence". Adding then "Independence" to the initial query would narrow down the results quite dramatically. In the same way, "Leonardo da Vinci" could yield several works and topics, among which "drawings" which could readily be used to precise the search intent. The same technique can be used to disambiguate a query, like adding “Elon Musk” to “Tesla”, or to look for unexpected conceptual intersections, like “Tesla” and “Drawings”.

The nature of entities is such that nothing prevents querying using a person, a location or even a document. A query consisting of an academic article would yield adjacent entities, e.g., authors, keywords and publishing details. Such queries can consist of a single entity, as well as two or much more. We refer to that technique as entity picking.

This section reports on the design and evaluation of ExplorationWall (EW), a system that implements entity picking.

The reliance of information exploration activities on search and iterative querying [Capra et al., 2007] tends to exacerbate any challenge in the basic search process. First, activities considered as related to exploratory search usually take place in areas that are unfamiliar to the user which makes query formulation intrinsically difficult. Second, the conventional search process relies heavily on text-based and window-based interaction, which are notorious weaknesses of touch devices, e.g., tablets and smartphones, on which text manipulation is made difficult by the absence of a physical keyboard, hotkeys or shortcuts, and the lack of an accurate selection tool [Esenther, 2006, Varcholik et al., 2012].

Therefore, we decided to take this challenge as an opportunity to develop a system specifically designed for touch devices, by using entities and developing affordances that would enable direct manipulation of entities for querying. Entity-based querying not only provides a substitute for typing, but also provides possible search directions to the user, who does not need to recall search terms and formulate queries, but can readily use entity search results to express and refine her search intents.

Three principles were derived from these requirements for the design of the system:

1. Querying and organization of information was to be done through direct manipulation of entities to provide a substitute for text entry and provide users with search directions in unfamiliar areas.

2. Results would consist of entities of different types, e.g. persons, topics, documents — by opposition to conventional purely document-based search results — each ready to be used as a new query or to refine or precise an existing query.

3. Multiple search sessions — or result sets — would share a common workspace to foster insights and parallel search, and address the limitations of windowed parallel search sessions.

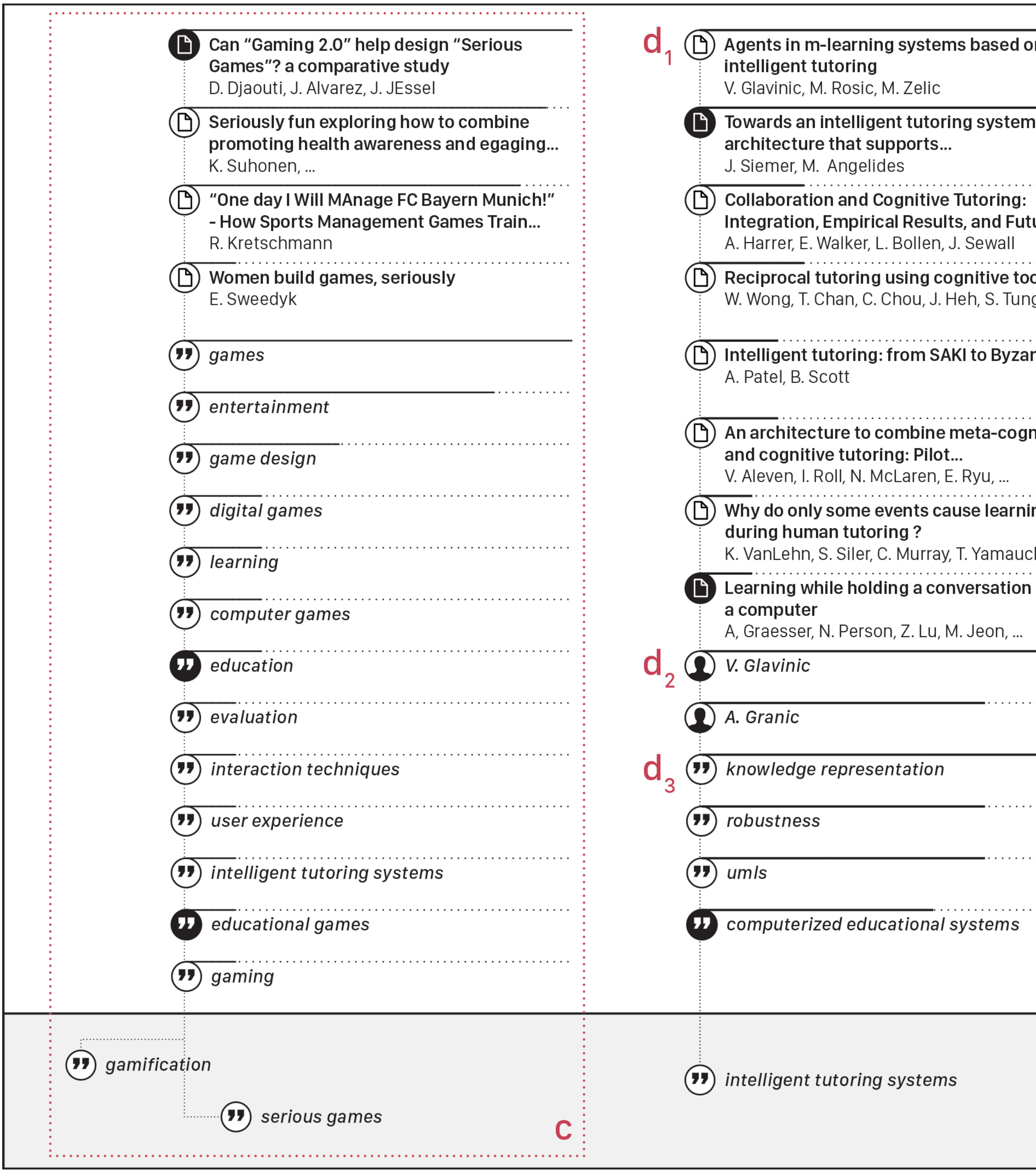
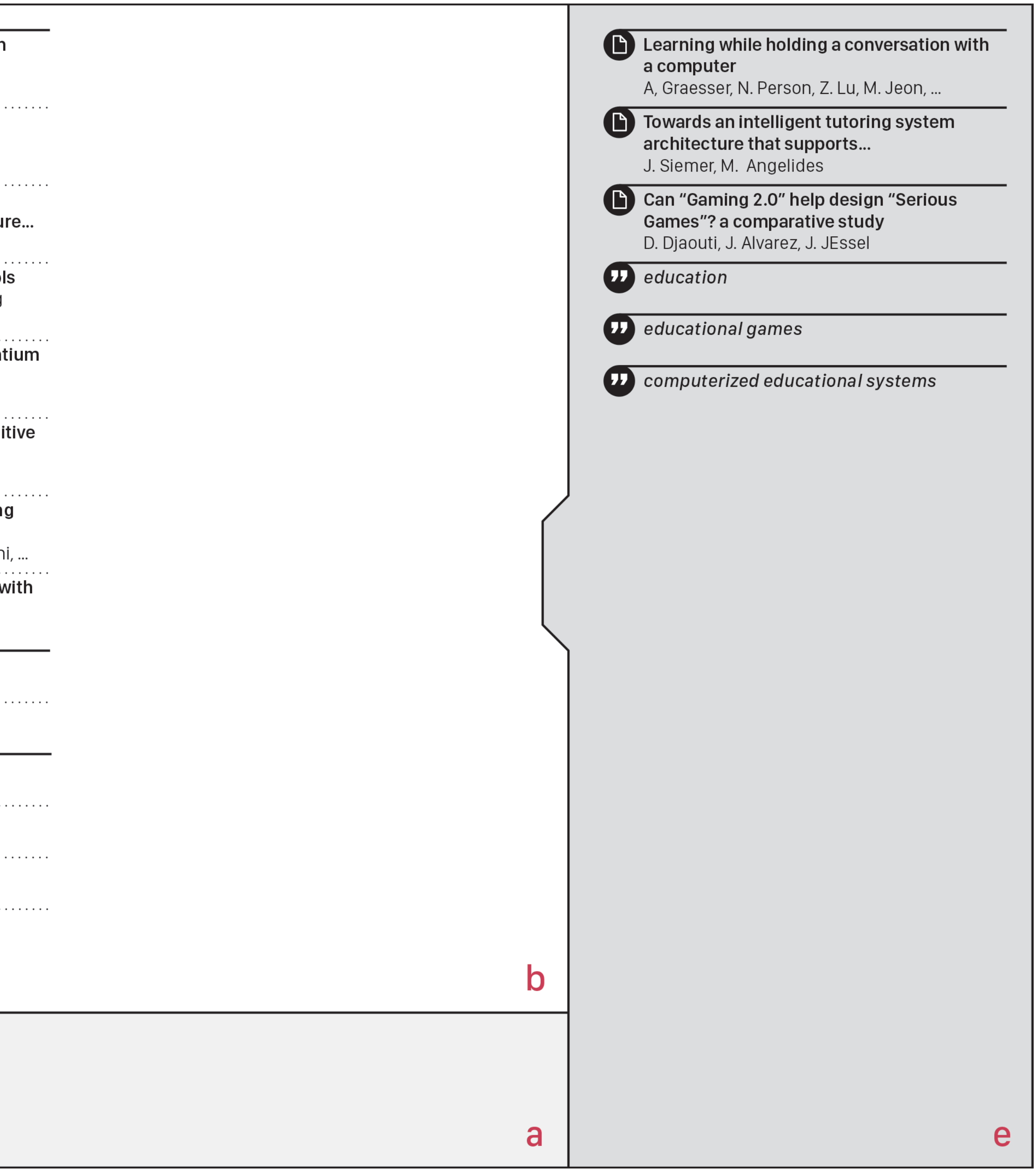


Figure 4.1. The ExplorationWall interface consists of the query area (a), the results area (b), search streams (c), 3 types of information items: papers (d1), authors (d2) and keywords (d3), and the reading-list drawer (e). Entities can be moved freely to the query area. A relevance gauge over their label visualizes their computed relatedness to the query. Queries consists of entity clusters positioned manually. The whole workspace is scrollable and horizontally unlimited.



Multi-touch gestures allows user to easily add or remove space between streams. Papers can be consulted by tapping on their icon.

4.1.1 Overview of the System

The system works as a full-screen standalone application.

The User Interface

The interface of ExplorationWall (Figure 4.1) consists primarily of the main workspace, which is divided into two areas, the query area at the bottom (Figure 4.1 a), and the result area on top (Figure 4.1b). An entity dropped in the query area constitutes a query and automatically yields a set of related entities (Figure 4.1 d) aligned vertically on top of it in the result area. We refer to this vertical presentation as a search stream (Figure 4.1 c). Any entity from the result set can be in turn drag-and-dropped to the query area to constitute a query and yield a new crop of related entities in a parallel search stream. Entities in the query area are freely positioned by the user. When two entities are put horizontally close enough together in the query area, they attach through a thin line that signifies that they will be considered as a single query. Their respective result set merge into a new one, now forming a single search stream containing a result set of entities that, if possible, relate simultaneously to both entities in the query. Queries can consist of a virtually unlimited number of entities. In the same way, there is no fixed limit to the number of parallel search streams that can be created in the workspace, which can be scrolled horizontally to provide some space or to retrieve some previous search stream. Horizontal space can also be added to, or removed from, a specific location using a conventional pinch gesture, the same pinch gesture can also be used to dilate or contract space, to quickly improve legibility of an area become cramped with information. Any entity taken from one search stream can be used either as a new query or to refine the query of an existing search stream.

This instantiation of ExplorationWall is developed around a collection of academic publications, for such information is readily structured into a graph of entities, namely authors/persons (Figure 4.1 d2), keywords/topics (Figure 4.1 d3) and articles/documents (Figure 4.1 d1), which constitute the three types of entities that are available here. Each entity is represented by a pictogram and a name or title label. Result entities are also displayed with a relevance gauge that indicates their relatedness to the query as estimated by the system's algorithm. Document entities also display the author's list in addition to their title and below it. Document entities can be tapped to reveal additional metadata and an abstract.

A reading list is accessible as a collapsible drawer on the right side of the workspace (Figure 4.1 e) and can be swiped open or closed. Entities of interest can be dragged and dropped inside it for storage and later use. Stored entities will appear highlighted in the workspace.

Interaction Scenario

On the 5th of August 2014, Alice has just heard about the ESA Rosetta space mission which probe had just reached the vicinity of the comet Churyumov–Gerasimenko. Having learned that the mission has launched ten years prior, using ExplorationWall, she is curious to see what kind of information she can find within the academic data of the system. She starts by instantiating the keyword “Rosetta” and uses it as a query. Without surprise, most of the results refer to the rosetta stone, she unfortunately does not see information related to the space mission. She adds the keyword “Comet” to the query, now all information is related to her subject of interest. She sees a number of articles, researchers and recommended topics on the subject. One article is about MIRO, a microwave instrument onboard the orbiter. She drops the article as a new query, and a parallel search stream opens. She sees the authors of the papers, related papers, and a few recommended topics, including “Instruments”. She adds that entity to her first query now consisting of “Rosetta”, “Comet” and “Instruments”. The refreshed results now offer a catalog of all instruments on board the spaceship to be inspected. Simply following her curiosity and without prior technical knowledge of the subject, Alice was quickly able to find new information of interest in a highly technical area, which would require higher cognitive load using conventional search tools[1] that forces the user to iteratively come up and formulate her search intent.

4.1.2 Overview of the Study

The main purpose of the evaluation was to observe the effects and implications of entity-based querying on search performance and search behavior. Therefore, ExplorationWall was compared to a search interface that was implemented following the interface principles of traditional search tools as seen in Figure 4.2.

The evaluation consisted of two tasks, a short one and a long one. We

[1] As conventional search tools would rely on her recollection of search terms of interest instead of providing search directions. Recognition has been shown to be less cognitively taxing than recall [Hearst, 2006].

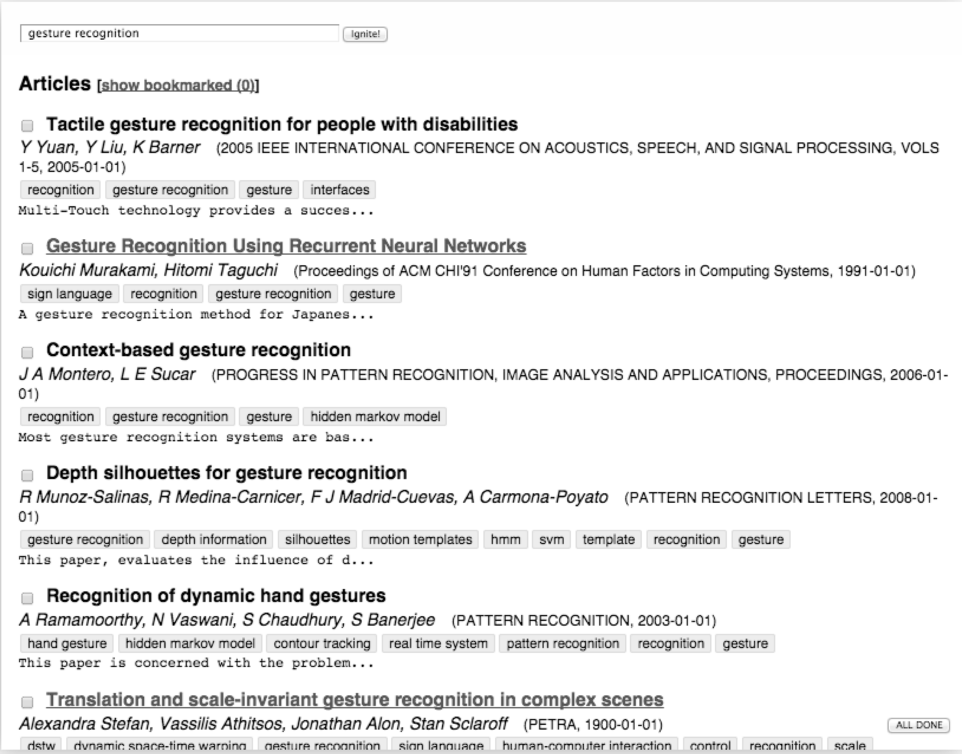


Figure 4.2. A screenshot of the baseline system replicating a conventional search interface with typed query.

chose 6 possible different topics for the two tasks: *crowdsourcing*, *smartphones energy efficiency*, *diagrams*, *semantic web*, *lie detection* and *digital audio effects*. In order to ensure that participants were not experts in the topics and could perform a real exploratory search, they pre-rated their familiarity with the topics on a 1 (less familiar) to 5 (most familiar) scale. The four less familiar topics were used in the tasks. Both tasks were performed with different topics, so the participants did not know the results from the previous task. For the short task, participants were given five minutes to address the following instruction: "Search and list 5 relevant authors, documents and keywords that you consider relevant in topic Y." For the long task, they were given twenty minutes: "Imagine that you are writing a scientific essay on the topic X. Search and collect as many relevant scientific documents as possible that you find useful for this essay. During the task, please, list what you think are the top five key technologies, persons, documents and research areas and write five bullet lines, which would work as the core content of the essay."

In this study, we followed a within-subjects experiment design[2] , counter-balanced by changing the order of the two tested interfaces, as well as the order of the two tasks. We recruited ten participants who performed the task using an iPad tablet. Before starting the main tasks, users received detailed instructions on how to use the interface and performed a five-minute training task on each interface. For text entry, we relied on the native virtual keyboard of the tablet. At the end of the sessions, participants were asked to answer the UES and SUS questionnaires for each interface.

We measured effectiveness, in other words, the quality of the information retrieved and displayed by the system, by computing usual metrics in information retrieval that are *Trecision*, i.e., found relevant information over all found information; *Recall*, i.e., relevant information found over all available relevant information; and *F-measure*, i.e., a single measure that combines both Recall and Trecision[3] In order to understand and compare users' search behavior, we analyzed participants' search trails using a method resembling White's [White and Morris, 2007]. In a similar manner, we looked for descriptive statistics of the search trails by selecting six parameters relevant to both interfaces:

- Number of queries: the total number of queries that were submitted during each task on both interface.

- Number of text entries per query

- Number of revisits: The number of revisits to a query or stream consulted

[2]A *within-subjects* experiment means we relied on the same pool of participants for testing both experimental conditions, as opposed to *between-subjects*, which would imply distinct pools of participants for testing each condition. Within-subjects experiment designs require less overall participants while minimizing random noise introduced by individual personalities, however, they require some care to avoid participants to transfer acquired knowledge from the first tested condition to the next. To that end, we carefully balanced among participants the order of the tested conditions, as well as the provided topics to be explored.

[3]The F-measure is especially useful when comparing two conditions, as Precision and Recall are generally involved in a trade-off. A system that displays every available item from the data after a query would measure perfectly in Recall, as all relevant items would have been retrieved, but would be unusable since the Precision measure would be very low, as these items would be lost in a sea of irrelevant ones. While a system that returns only one relevant item would have perfect Precision but very low Recall. The F-measure therefore measures any benefit beyond such trade-off.

earlier in the current trail.

- Number of branches: The number of times a subject revisited a query or stream on the current trail and then proceeded with the formulation of a new query.

- Number of queries per minute: the number of queries per minute that were submitted during each task on both interface.

- Number of parallel queries: Number of parallel streams produced with ExplorationWall or number of tabs opened with the baseline.

We also measured usability and engagement through standard questionnaires: System Usability Scale (SUS) [Brooke et al., 1996] and the User Engagement Scale (UES) for exploratory search [O'brien et al., 2014]. SUS consists of a ten-item questionnaire and is a widely used and validated for measuring perceptions of usability. Since the degree of user engagement is a strong indicator of exploratory search performance [White and Roth, 2009], we chose to use UES for exploratory search, considering six different dimensions: Aesthetics, Focused Attention, Felt Involvement, Perceived Usability, Novelty and Endurability aspects of the experience.

4.1.3 Findings

Effectiveness: ExplorationWall shows substantial improvement in effectiveness in the long task. The improvement was found to hold for task-level measurement, but also for averaged interaction-level measurement for which the recall and the F-measure were found to be substantially higher compared to the baseline (Figure 4.3). These measures indicate that, on average, participants covered more ground over the given topic of interest using ExplorationWall when compared to the baseline, without sacrificing any precision or quality in the search results. No significant differences between the systems were found in the short task or in the expert ratings.

Search Trail Analysis: The users in the ExplorationWall condition were found to use all of the measured interaction features significantly more than the users in the baseline condition in the long task. Differences were also found in the short task. The users in the ExplorationWall condition typed less, branched more, and used more parallel queries.

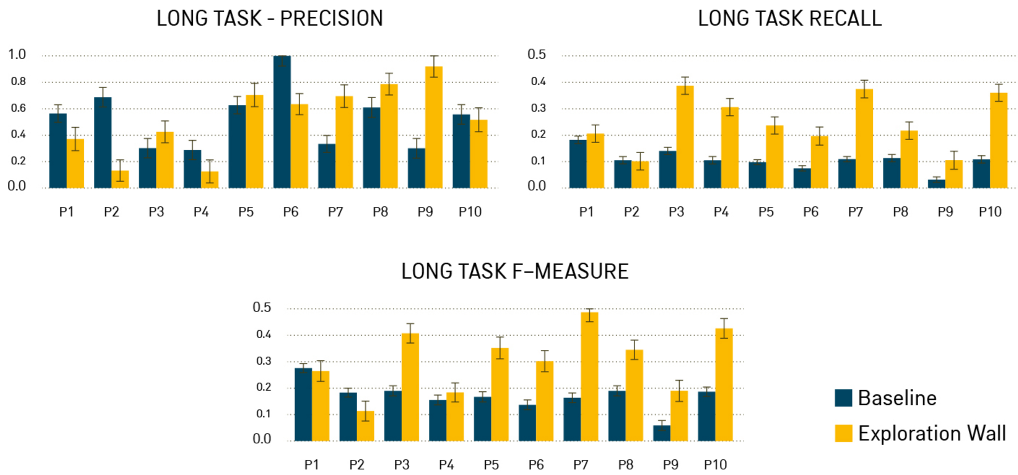


Figure 4.3. The effectiveness results showed a significant advantage for ExplorationWall in *Recall*, i.e., the amount of relevant documents that were found over all relevant documents available, while *Precision*, i.e., amount of found documents that were relevant over all found documents, showed non significant differences, resulting in an overall significantly better *F-measure*, i.e., the harmonic mean of precision and recall. Such results show that participants covered more of the explored topic using ExplorationWall.

Usability and Engagement: The results for the mean of answers of the SUS questionnaire showed a significant difference between the two systems, revealing higher usability for ExplorationWall. The results of the UES questionnaires are also favorable for ExplorationWall.

The difference in recall proves that more relevant documents were retrieved when using ExplorationWall, which can be explained by our measure of a more active search behavior that was induced by the use of ExplorationWall, with more queries per minute and more branching. Such more active behavior could in other cases be the result of a poor search performance, where unsatisfying search results force users to frequently reformulate their query. However, in the present case, the precision measure proves the quality of each result set was comparable to those obtained in the baseline.

Furthermore, participants voluntarily avoided to use text entry when using ExplorationWall, preferring the direct manipulation, which again, could in other cases be a sign of poor implementation, but the results from the UES questionnaire also show a better user engagement, a factor that is likely to have contributed to the more active search behavior, and validate the hypotheses that the design of ExplorationWall makes the direct manipulation of entities preferable to conventional text-based search mechanics.

4.1.4 Design Implications

Direct manipulation and selection of entities for querying proved successful, in both user preference and search performance. Allowing users to search for not only documents but various entities helped users make sense of the information space linked to an unfamiliar topic. **Such results validate the *entity picking* technique for entity search and provides a suitable alternative to text-based querying when designing future touch-based interfaces.**

The multiple-stream layout also proved key in fostering insights over multiple parallel search sessions as suggested by users' search behavior and overall better coverage of the explored topic. Users were able to organize and visualize past and new search efforts simultaneously and voluntarily took advantage of this. **These results indicate the importance of valuing search efforts instead of considering it as ephemeral, by facilitating storing, organization and visualization of found information.**

4.2 IntentStreams: Querying Through Parallel Search Intent Modelling

As explained in the Introduction, typed queries rely on a user's current knowledge, thus creating limited opportunity in exploratory settings for inputting accurate search intents, since users are often plunged in unfamiliar information environments. Allowing the system to receive relevance feedback from the user regarding the perceived quality of individual results to a query is a known method for improving the expression of a user's search intent, however the lack of possibility for inputting the reason for a any result's level of relevance requires the user to go provide a lot of feedback for it to be effective.

An enticing workaround would be to present the user with a representation of her search intent as perceived by the system, that she can in turn tweak so that it better matches her actual intent. Such an approach, commonly referred to as a *User Model* consists of creating a dynamic template of the user and infer in real-time her knowledge through her actions [Suchman, 2007]. Applied to search activities, this technique is referred to as *Interactive User Intent Modelling* by Ruotsalo, Głowacka, and colleagues, and has proven effective in exploratory search tasks

[Ruotsalo et al., 2013a, Glowacka et al., 2013].

User intent modeling takes an initial typed query to return a result set of documents. From this set, the system extracts a set of entities that are most representative of the set of retrieved documents, and weights these entities according to their estimated relevance. These entities represent the intent model, which is presented to the user who can modify individual weights, e.g., increase or decrease their relevance, or completely discard them. The system then refreshes the results to better reflect the updated intent model.

This technique offers a slightly different way of supporting query formulation compared to entity picking: Providing relevance feedback on a set of entities is more input heavy and could hinder the overall fluidity of the exploration. Also, by providing only entities extracted from the retrieved set, the technique creates limited opportunities for branching away from the initial topic.

For this reason, we were interested in applying this technique while supporting parallel sessions and support the transport of entities from one session to another. To that end, we decided to reuse the overall structure provided by the user interface of ExplorationWall in section 4.1 and adapt it for the new prototype, referred to as IntentStreams (IS).

The evaluation results of ExplorationWall showed that the parallel search streams configuration fostered a more active exploratory behavior, with participants being more aware of past search results as measured through the number of revisits and branching of the exploratory path. We were interested in further investigating such an effect on users' exploratory behavior. Since the focus was not specifically on supporting exploration on touch devices anymore, and to allow comparison to a baseline consisting of a naturalistic search setting, the system should support entity based interaction, but yield the same document-based results as the baseline.

Following is a summary of the main requirements for the design of IntentStreams:

1 The prototype should support direct manipulation of entities and retrieve document-based document set.

2. A user model was to be implemented in the form of a set of weighted keywords.

3. Querying was to be done through relevance feedback.

4. The prototype should support parallel search streams and support the transport of entities across search streams.

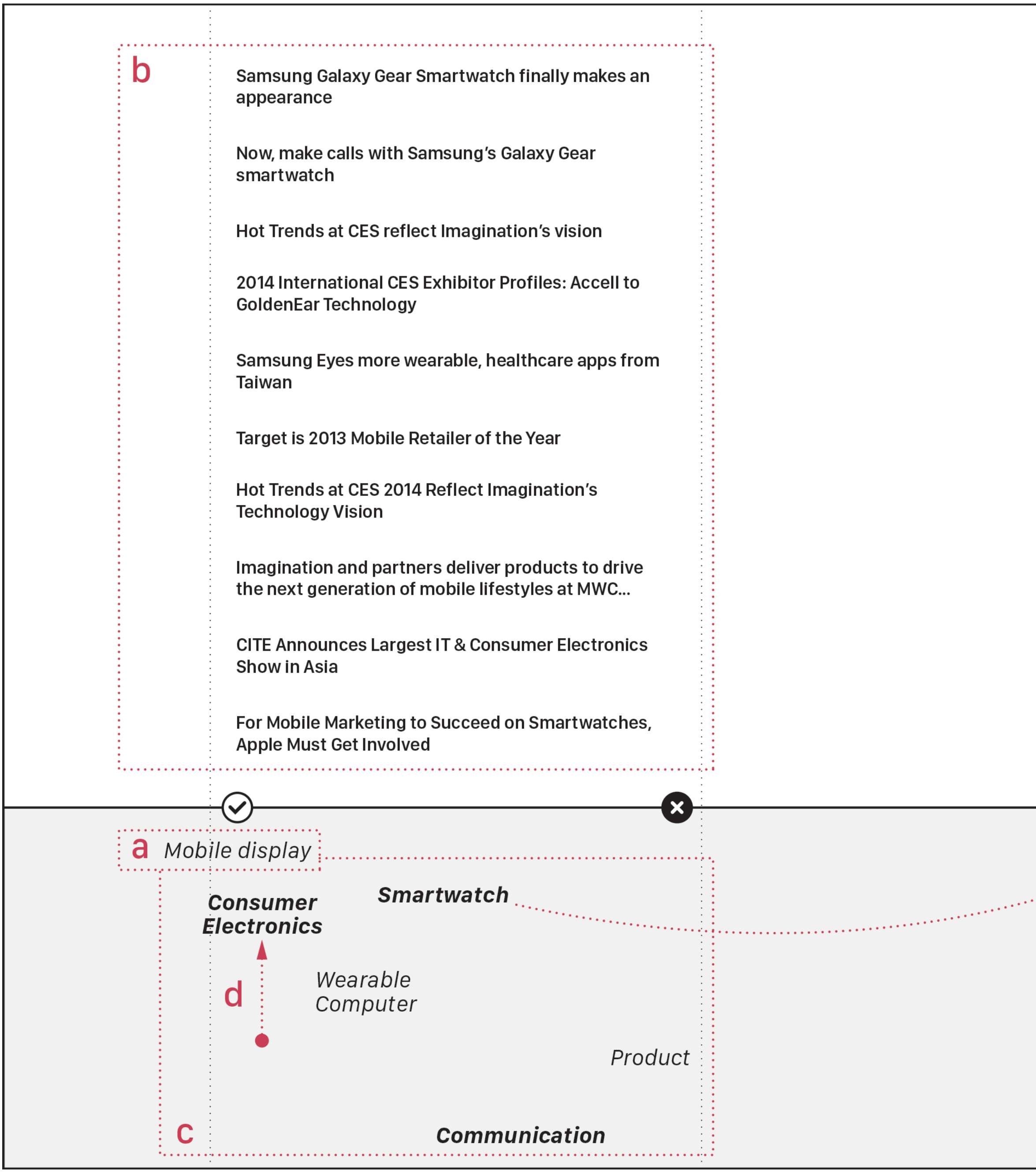
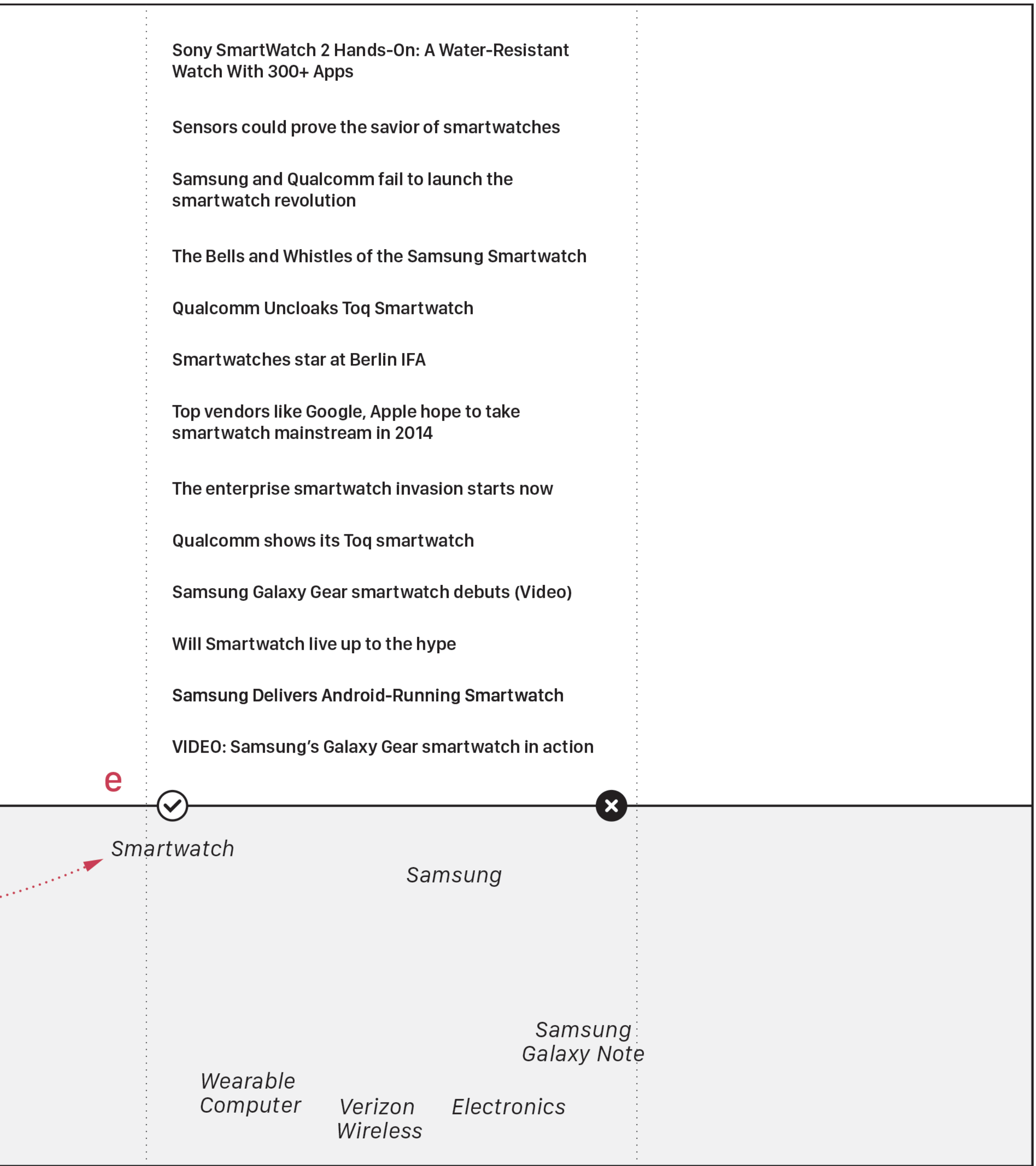


Figure 4.4. The user interface of IntentStreams. The first query, in this case *mobile display* (a), returns a search stream consisting of news articles most relevant to the query (b), as well as a set of most relevant keywords extracted from a larger set of related articles (c). These keywords represent the user model, or how the system perceives the user's search intent. The user can modify the weight of the keywords by sliding them vertically, thus emphasizing or reducing their desired importance (d), after which the stream will refresh, updating articles and keywords accordingly.



If dropped outside their initial stream, keywords will either trigger a new search stream (e) or be passed to an already existing parallel stream. Articles can be consulted by clicking on their title, which opens a reading panel.

4.2.1 Overview of the System

We refer to the present prototype system as IntentStreams. In the instance of the prototype that was used for evaluation, it was connected to a repository of English language editorial news articles crawled from publicly available news sources. The database contained more than 25 million documents.

The documents were originally collected for monitoring media presence of numerous interested parties, hence the collection's wide variety of topics.

The User Interface

IntentStreams provides a workspace divided into two areas: the keyword area at the bottom and the document area on top (Figure 4.4). By clicking on the workspace, the user is prompted to type an initial query (Figure 4.4 a). That query yields two sets of elements: A ranked set of news articles displayed through their titles is listed vertically in the document area (Figure 4.4 b), while below, in the keyword area, the typed query has split into a variety of keywords, i.e., the user intent model (Figure 4.4 c).

These keywords, between five and ten[4], are separated horizontally for legibility, and positioned vertically according to their estimated relevance, the higher the more relevant. We refer to such a vertical arrangement of documents and keywords as a search stream.

Clicking, on a news article title will reveal its content on top of the result list. Click-and-holding a news article will highlight in the intent model the keywords directly related to the article (rarely will a found article relate to all keywords in the intent model simultaneously). In a similar fashion, click-and-holding a keyword will highlight related articles.

Keywords can be moved vertically to adjust their desired level of relevance (Figure 4.4 d). Keywords that have been manipulated will change color, and a validation button will appear for triggering the update of the results, which allows the user to adjust several keywords before triggering any change.

Additional search streams can be created by clicking on the workspace, anywhere outside of an already existing stream. This allows a user to compare the results in parallel search sessions. Parallel search streams

[4] The user intent model shows only extracted keywords whose estimated relevance is above an arbitrary threshold, set through trial and error, hence the variable amount of displayed keywords. Therefore only the most representative keywords are displayed to limit overwhelming visual clutter.

each exhibit their own set of articles and user intent model. Any keyword from one intent model can be transferred to a parallel one. A keyword can also be dragged outside of an existing search stream, which will trigger the creation of a new stream (Figure 4.4 e).

The workspace can be scrolled horizontally to make space for additional search streams and not be limited by the boundaries of the display, with the effect that after creating a sequence of search streams, the horizontal layout will often reveal the search history of the session. Search streams can also be dragged and rearranged horizontally.

4.2.2 Overview of the Study

We evaluated the system to find out if and how IntentStreams supports parallel browsing and branching behavior. IntentStreams was compared against a baseline system with an interface similar to a traditional Google search interface (see Figure 4.2 as it was the same baseline system we used for evaluating ExplorationWall in section 4.1). Our hypothesis was that, compared to the baseline, IntentStreams generates (1.) more parallel streams, (2.) more revisits, and (3.) more branches. We used the following metrics: the number of parallel streams, number of revisits, and number of branches.

In the baseline, the number of parallel streams denotes the number of tabs opened, a revisit indicates returning to an already open tab, and a branch denotes a query updated after a revisit. In IntentStreams, a revisit occurs when a user performs certain activities (opening an article, weight change) on a previously created stream. A branch occurs when a new stream is created from an existing one. That includes both creating a new query by dragging a keyword or updating the existing stream by modifying the weights of its keywords.

We evaluated the system with thirteen volunteers. We used a within-subject design, where participants were asked to perform two tasks, one with IntentStreams and one with the baseline, after each receiving detailed instructions and having performed a five-minute training session. The task was formulated as follows:

You have to write an essay on recent developments of X where you have to cover as many subtopics as possible. You have 20 minutes to collect the material that will provide inspiration for your essay. You have 5 additional minutes to write your essay.

The two tasks performed by the participants covered two topics: (1)

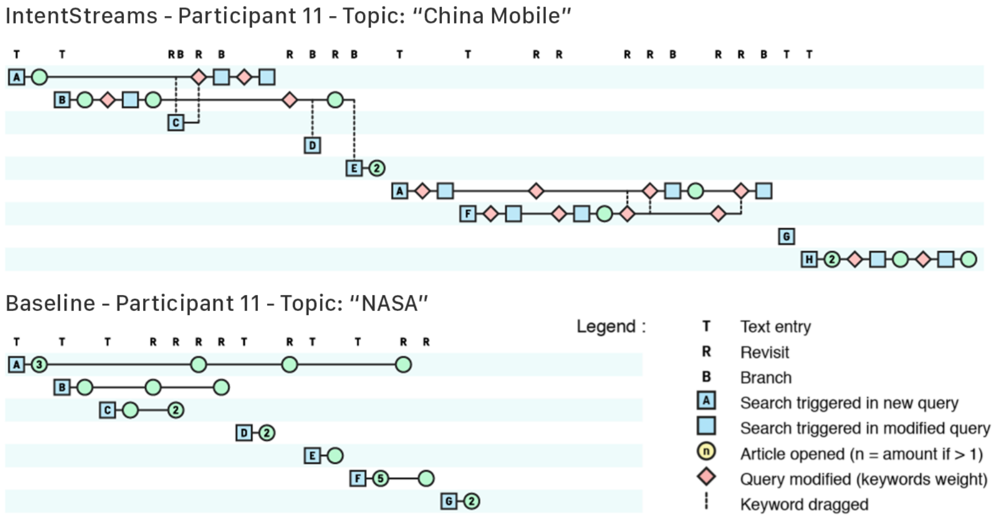


Figure 4.5. Visualization of two search trails resulting from the same participant during a similar amount of time, once performed using IntentStreams (top), once using the baseline (bottom). Rows visualize parallel searches while columns show the user's sequence of actions. This example illustrates the difference in revisiting and branching behavior fostered by each system, with IntentStreams leading to a much more active exploratory behavior.

NASA, and (2) *China Mobile*.

To evaluate the system, we connected it to a news repository of English language editorial news articles crawled from publicly available news sources from September 2013 to March 2014. The database contains more than 25 million documents. The documents were originally collected for monitoring media presence of numerous interested parties, and hence the collection has wide topical coverage.

The baseline system was connected to the same news repository as IntentStreams. In the baseline system, users could type queries and receive a list of relevant news articles. To start a new parallel query, a new tab had to be opened.

4.2.3 Findings

IntentStreams on average generated 7.84 more queries (SD = 7.27), 6.38 more parallel streams (SD = 4.03), 4.54 more revisits (SD = 4.52), and 3.62 more branches (SD = 4.01). A paired t-test indicated that all those differences were statistically significant ($p < 0.01$).

Results show that users created more parallel streams than opened new tabs. While the system allows the creation of parallel streams, the users revisit earlier ones consistently.

With IntentStreams, more queries and parallel streams were created through branching as seen in Figure 4.5. A visual representation of a participant's search behavior shows the difference between the linear search behavior in the baseline and the more articulated search behavior in IntentStreams. Further, IntentStreams has shown to better support exploration, as can be seen from the higher number of queries.

4.2.4 Design Implications

IntentStreams is an example of how to successfully utilize intelligent agents, in this case, user intent modeling, in a **user-driven way that puts users in control** and allow them to tinker and explore the results. Direct manipulation and selection of entities for querying proved successful again, and the study confirmed **the benefits of having multiple-stream layout in fostering insights** over multiple parallel search sessions as suggested by users' as the search trails proved richer and less linear, demonstrating better overall coverage of the explored topic by participants.

4.3 Contribution

The aim of this chapter was to answer RQ1: *How can entity-based querying benefit information exploration?* User experiments with ExplorationWall yielded substantial results, demonstrating that entity picking could rival and even outperform conventional text-based querying in search performance, while having a positive effect on search activity, behavior and engagement. With IntentStreams, we investigated an alternative technique through user intent modeling and relevance feedback, more input heavy, while further reducing differences between the prototype and a naturalistic baseline condition and we still obtained significantly positive results regarding search performance and engagement. Both these results comforted our confidence in the potential of entity-based affordances for information exploration and brought satisfactory answers to RQ1.

5. Enabling Orientation in the Information Space

Exploratory activities rely on search and discovery, but in a larger context, in which knowledge and intents keep evolving. Information exploration by definition describes open-ended and potentially long-term activities. As such, the temporal and spatial components of exploration are usually provided with little support in our information practices, because of tools that tend to consider search as an ephemeral single-user activity.

Orientation – the knowledge of one's location in relation to one's surroundings – is necessary in making a conscious decision of where to go next, which makes it a crucial ability in getting the most benefit from exploration. Unfortunately, orientation is made difficult by search tools returning narrow slivers of the information space, ordered according to hidden criteria.

With our larger goal in mind – translating properties of spatial exploration into the information space – we were interested in designing and investigating visualizations that would provide users with a map of such space.

The biggest challenge lies in the high multi-dimensionality of any information space: the semantic distance between elements depends on so many criteria that any attempt of absolute translation into two or even three dimensions suitable for visualization purposes, would be highly inconsistent and unusable.

The available solution was to let the user determine criteria relevant to her needs or interests so that the mapped information space is specific to a user and a given context, e.g., a given project or time. Entities and entity-based querying seem convenient in not only expressing a search intent but also letting a user define and demarcate an information space that is relevant to her interest, following a number of criteria that make it manageable to visualize and explore.

A resulting limitation of such an approach is that enabling orientation for one given user in one given context only partially addresses the challenge at hand. Therefore, a remaining challenge was to investigate techniques for users to share such information spaces and find common ground.

This chapter reports on the design of two prototype systems designed to support defining coherent information spaces through entity affordances. The first one, summarized in Section 5.1 and reported in Publication III, implements information mapping for exploration and allows investigation of its benefits. The second one, summarized in Section 5.2 and reported in Publication IV, allowed us to investigate the possibility of finding common ground in collaborative exploration through entity-based affordances. Both systems and their user-based evaluation brought satisfying answers to RQ2: *How to demarcate and visualize a coherent information space through entity-based affordances?*.

5.1 RelevanceMap: Multidimensional Perception of the Information Space

Chapter 1 has addressed the limitations of one-dimensional ranked lists of results and how they fail at providing a meaningful overview of the result space. Multiple simultaneous criteria allows for rich mapping of the result space, while at the same time allows the user to demarcate the information space through entity-based querying. By limiting the mapping to user-defined criteria, the mapped information space becomes a subset of the available data set. Remembering the *Information Flaneur* [Dörk et al., 2011], the desired resulting information map should be reminiscent of a city, with boundaries as gates, consisting of user-defined criteria; more or less semantically consistent or diverse areas as districts; and opportunities to constitute parallel sets of criteria, like paths to neighboring cities.

To that end, the design or RelevanceMap had to follow the following requirements:

1. Possibility for a user to arrange criteria in the form of manipulable entities on a map.

2. Have the system return all available relevant information.

3. Have retrieved information visually encoded on the map as a topology, with respect to the user-defined criteria.

4. Allowing through interaction with the map, the progressive exploration

of chosen areas of the map by revealing location-relevant information content.

5.1.1 Overview of the System

The User Interface

The user interface of RelevanceMap (Figure 5.1) consists primarily of a map area (Figure 5.1 a), with a companion panel on the right for displaying search results (Figure 5.1 b). The map area contains a text input field (Figure 5.1 c). Typing a phrase in the input field adds it as an interactive entity (Figure 5.1 d) on the map that can be freely repositioned. Repeating this process, a user can input multiple query phrases describing topics or areas of interest, and arrange them on the map. After each phrase input, the system retrieves from the data set all documents that relate to one or several of the query phrases. The number of retrieved documents being possibly quite large, possibly in the thousands or more, it is not possible to inspect them all at once. Therefore, these documents are instead displayed on the map as a visual marker.

Document markers each take the form of a semi-opaque dark circle with a distinct location and radius (Figure 5.1 e). The location of a document marker depends on the individual relevance estimation of the document to each of the query phrases. The positioning of a document marker can be described by the effect of springs, that would link each query phrase to the document marker (Figure 5.1 f). The higher the relevance estimation of the document to the query phrase, the stronger the spring. As a result, the more semantically close a document is to a query phrase, the more spatially close its marker gets.

Since the end location reveals only relative relevance, i.e. a document with nothing but very weak relevance scores to each of the query phrases can share the same position as a document with strong relevance scores, we have introduced the notion of overall relevance, which is the sum of all the relevance scores of a document. This value is then encoded as the radius of the document marker.

The transparency and variable radius of all document markers together contribute to building a topography of how a document set is populated with respect to a user's interests. The superimposition of document markers creates darker areas informing the user of a high density of information, while light and white areas indicate a gap in the semantic distribution.

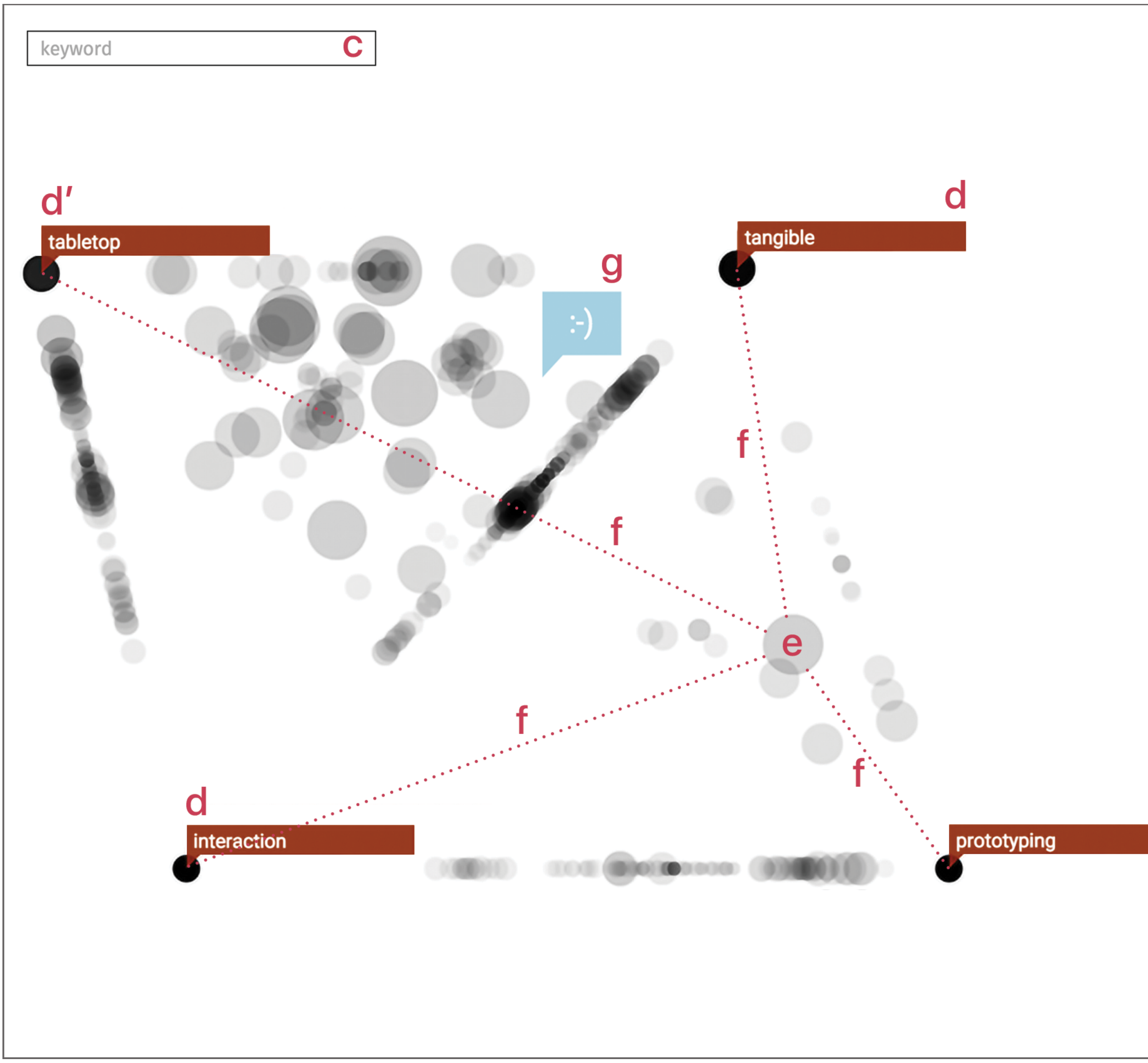
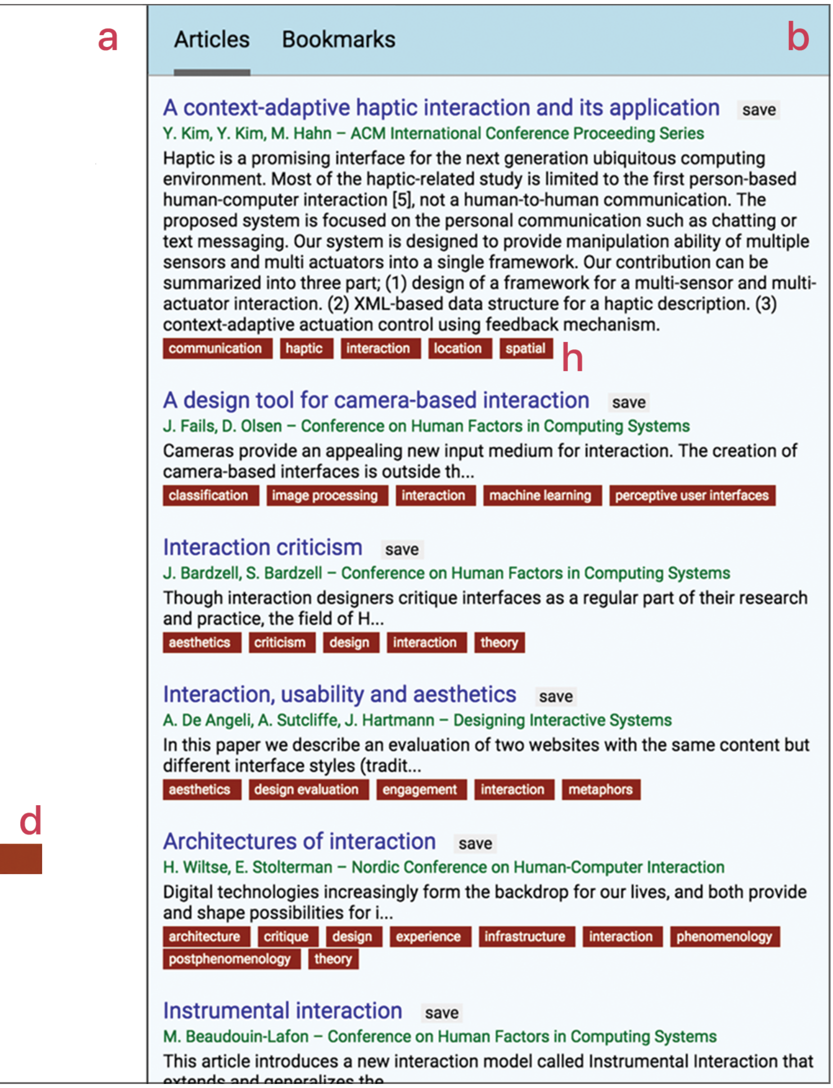


Figure 5.1. The relevance map displays documents in relation to multiple entities, displayed as red text labels (d). The exploration cursor (g) is located at the user-specified position to be used for the re-ranking of all related documents in the right panel (b).


a
Articles
Bookmarks
b
A context-adaptive haptic interaction and its application
save
Y. Kim, Y. Kim, M. Hahn – ACM International Conference Proceeding Series
Haptic is a promising interface for the next generation ubiquitous computing environment. Most of the haptic-related study is limited to the first person-based human-computer interaction [5], not a human-to-human communication. The proposed system is focused on the personal communication such as chatting or text messaging. Our system is designed to provide manipulation ability of multiple sensors and multi actuators into a single framework. Our contribution can be summarized into three part; (1) design of a framework for a multi-sensor and multi-actuator interaction. (2) XML-based data structure for a haptic description. (3) context-adaptive actuation control using feedback mechanism.
communication
haptic
interaction
location
spatial
h
A design tool for camera-based interaction
save
J. Fails, D. Olsen – Conference on Human Factors in Computing Systems
Cameras provide an appealing new input medium for interaction. The creation of camera-based interfaces is outside th...
classification
image processing
interaction
machine learning
perceptive user interfaces
Interaction criticism
save
J. Bardzell, S. Bardzell – Conference on Human Factors in Computing Systems
Though interaction designers critique interfaces as a regular part of their research and practice, the field of H...
aesthetics
criticism
design
interaction
theory
Interaction, usability and aesthetics
save
A. De Angeli, A. Sutcliffe, J. Hartmann – Designing Interactive Systems
In this paper we describe an evaluation of two websites with the same content but different interface styles (tradit...
aesthetics
design evaluation
engagement
interaction
metaphors
d
Architectures of interaction
save
H. Wiltse, E. Stolterman – Nordic Conference on Human-Computer Interaction
Digital technologies increasingly form the backdrop for our lives, and both provide and shape possibilities for i...
architecture
critique
design
experience
infrastructure
interaction
phenomenology
postphenomenology
theory
Instrumental interaction
save
M. Beaudouin-Lafon – Conference on Human Factors in Computing Systems
This article introduces a new interaction model called Instrumental Interaction that

Initially, the result list displays all related documents in a long list, ranked according to the overall relevance of each document. Exploring the content of the map and accessing documents of choice requires the user to tap somewhere on the map, to position a cursor (Figure 5.1 g) inspired by the one used in Google Maps. This cursor tells the system what area of the map the user is interested in discovering. As a result, the system re-ranks the list of all related documents in a way that prioritizes the ones whose markers are in the vicinity of the cursor. Details of the methods used to compute the visualization and re-ranking are provided in Publication III [Klouche et al., 2017].

The actual method through which the ranked list is computed takes into account both the marker-to-pointer distance and the overall relevance value of each document. The system then re-ranks the complete list of retrieved documents according to these values.

The pointer can be re-positioned by dragging or tapping on the map. Any change in the pointer position or query marker organization triggers a re-ranking of documents based on their overall relevance and proximity to the pointer.

The ranked articles appear in a conventional one-dimensional scrollable list layout in the result panel, with title, authors and publishing information, abstract and keywords. Keywords from the results panel are readily usable as interactive entities and can be dragged to the map to become new query phrases (Figure 5.1 h).

5.1.2 Overview of the Study

Twenty participants took part in a controlled laboratory experiment in which RelevanceMap was compared to a conventional ranked list visualization in two basic tasks: perception and retrieval.

The perception task sought the understanding of the benefits of the visualization in perceiving the distribution and density of resulting documents with respect to multiple query phrases. The retrieval task sought the understanding of the benefits of the visualization in re-ranking the results according to a user-specified distribution over the importance of the different query phrases. The benefits were measured with respect to task completion time and effectiveness (quality of the perception or retrieval).

The perception task aimed to measure task completion time and effectiveness, to help understand how a document space is populated and organized with respect to specific query topics. Participants were asked the two

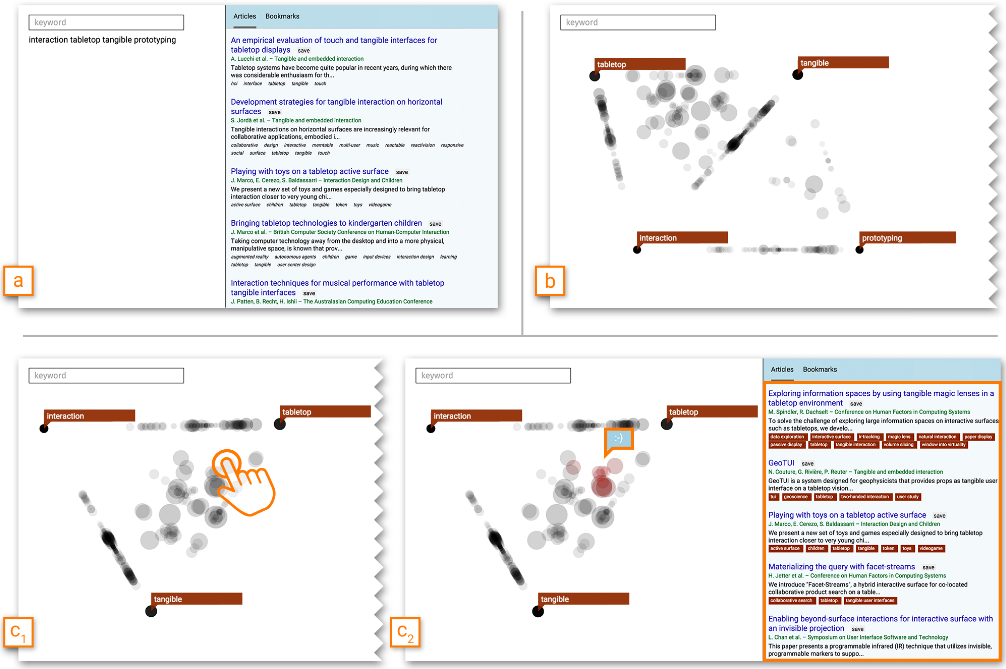


Figure 5.2. In the perception task, participants must identify the two and three keywords out of four that are the most related to relevant information. In the baseline (a), they must skim through the ranked list of results to infer the most prevalent keywords from the top articles. Using the relevance map, they must interpret the distribution of document markers. In the retrieval task, participants must find an article that shows a high relevance to one keyword (say, tabletop), and a lesser relevance with two other keywords (say, tangible and interaction). Using the baseline (a), they must query the three keywords, then find a fitting article in the result list. Using the relevance map, they point (by tapping on the touch-enabled monitor) at an area between the three keywords (c1), somewhere closer to tabletop than tangible or interaction, which triggers a re-ranking of retrieved articles based on the selected position (c2). The participant should be able to select one of the top articles as a fitting task outcome.

following questions: (1) "Out of the 4 topics provided, which 2 topics are related to the highest amount of relevant documents?", and (2) "Out of the 4 topics provided, which 3 topics are related to the highest number of relevant documents?" (Figure 5.2).

The retrieval task aimed to measure task completion time and effectiveness in finding documents with varying multi-dimensional relevance toward several topics. Participants were given the following instruction: "Find one article that is highly relevant to *Topic A* and slightly related to *Topic B* and *Topic C*.". The task was then repeated one more time with a different topic priority: "Find one paper that is highly relevant to *Topic B* and slightly related to *Topic A* and *Topic C*."

5.1.3 Findings

The results of the experiments show significant improvements in task completion time in both perception and retrieval, without compromising effectiveness.

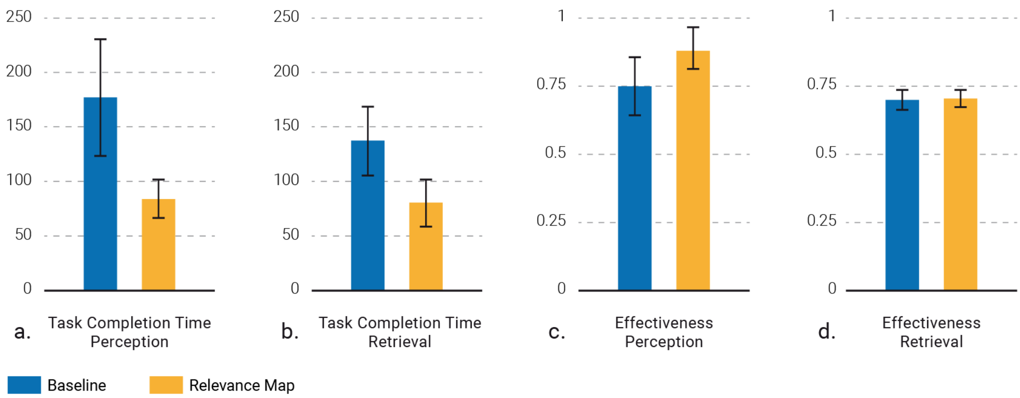


Figure 5.3. Results from the performance measures displayed for both systems with confidence intervals for: (a) task completion time in the perception task and (b) task completion time in the retrieval task with the mean duration (lower is better), (c) effectiveness in the perception task with the mean topic quality, and (d) effectiveness in the retrieval task with the mean document quality (higher is better).

In the perception task, participants were able to use the relevance map visualization to make decisions with greater accuracy and 111% faster. The visualization allowed the participants to understand more accurately the distribution of information with respects to the multiple aspects of the query.

In the retrieval task, documents fitting complex criteria were retrieved 70% faster using re-ranking through interaction with the relevance map.

While finding documents with different relevance to several topics requires users to go through long lists of results and assess the relevance of individual documents, our proposed method for re-ranking through pointing at the map successfully narrows down the top results to documents that fit the criteria.

The quality of the task outcome was the same in both conditions in the retrieval task. A possible reason for equal performance is the absence of strict time constraints for participants to complete the tasks. It is possible that a constrained time to complete the task would have negatively impacted the quality of the task outcome for the baseline, as the participants would not have been able to carefully examine the list to find a fitting article, but would have been forced to skim, resulting in possibly lower quality of selected topics and articles.

5.1.4 Design Implications

While the relevance map utilizes a similar result list component as the baseline system, the visualization, and the possibility to quickly re-rank the results help avoiding the trap of the ten blue links. The simplicity of the search engine is here substituted by a necessarily more complex behavior, but users showed an encouraging adaptability and enthusiasm to the new paradigm. This suggests that **richer user interfaces for accessing information can be accepted and worth being explored**.

5.2 QueryTogether: Collaborative Orientation

Search is often thought of as a solitary user activity, focusing on eliciting a user's information needs and improving search-result relevance. Recently, increasing attention has been devoted to search as a collaborative activity that is often co-located, spontaneous and initiated informally from a dialogue [Brown et al., 2015, Morris et al., 2010]. Users are inspired or informed by others' searches and can distribute search efforts, exploring the information space in parallel. Despite the increasing number of situations in which several co-located people engage in collaborative search, available devices, and public screens are not effectively used for synchronous collaboration.

Section 5.1 makes the case for dynamic mapping of an information space to make it relevant to a specific context and user. In the case of collaboration, this raises the question of how entity-based techniques for demarcating an information space of interest can be used effectively among multiple users. To investigate the question, We wanted to focus on a common ubiquitous computing scenario in which several co-located users spontaneously engage in collaborative search, using personal devices as well as available large screens or projectors. In particular, we investigate how entity-based user interfaces and search systems can facilitate collaboration across such devices.

5.2.1 Design Rationale

- *Entity-centric exploration*. The units of search and collaboration can be any information entities – such as documents, keywords, and authors – that can be shared for collaboration or used as queries to trigger exploration. The design should also provide different starting points for exploration, including not only entities suggested by peers but also entities suggested by the system. System suggestions should be provided every time a new query is triggered, so users are always provided with possible directions for future exploration. In both cases the suggested entities should be encapsulated in *interactive search objects* that can be directly used to trigger new queries and explore new directions.

- *Flexible use of devices*. To study the effect of entity-centric exploration in a scenario that reflects the current trends as closely as possible, one design goal is to make the system usable from a variety of devices and thus support different modalities (e.g., mouse and touch) and different platforms. To facilitate interaction on smaller devices (and on touch screens in general), the main features such as querying and sharing should not necessarily require typing. Enabling typing-free interactions may also prevent unnecessary overhead when accepting system suggestions and thus lead to less distraction and better exploration.

- *Support for diverse working styles*. Previous work on co-located collaboration has stressed the importance of supporting a variety of working styles ranging from individual work to tight collaboration [Scott et al., 2004, Tang and Joiner, 2006]. It is important to allow for various degrees of coupling as at times, the work is more efficient if it is performed by an individual or loosely coupled. This is also important due to the fact that in some instances, it might be appropriate to allow for maintaining the privacy of the information being manipulated by participants [Stefik et al., 1987]. Moreover, the system should allow for flexible switching between different working styles. Users should have the option to work independently and to decide if and when to share and whether to share privately or publicly. While this could lead to more effective collaboration [González-Ibáñez et al., 2013], it could also be beneficial from a privacy perspective, as users decide what to share without needing to disclose their entire search log.

5.2.2 Overview of the System

The User Interface

Since QueryTogether adapts the user interface from ExplorationWall, this section will focus on describing the specific features enabling collaborative use. Please refer to subsection 4.1.1, or to Publication IV [Andolina et al., 2018a] for the full description of the interface.

As seen on Figure 5.4, The side panel/reading list, already present in ExplorationWall, now additionally displays at the bottom the user list, where each user's name is displayed as a label along with his or her share status, i.e., "public" for shared devices and common workspaces, or "private" for users with individual/private devices. In the prototype implementation, the status was to be selected at login, along with a user name.

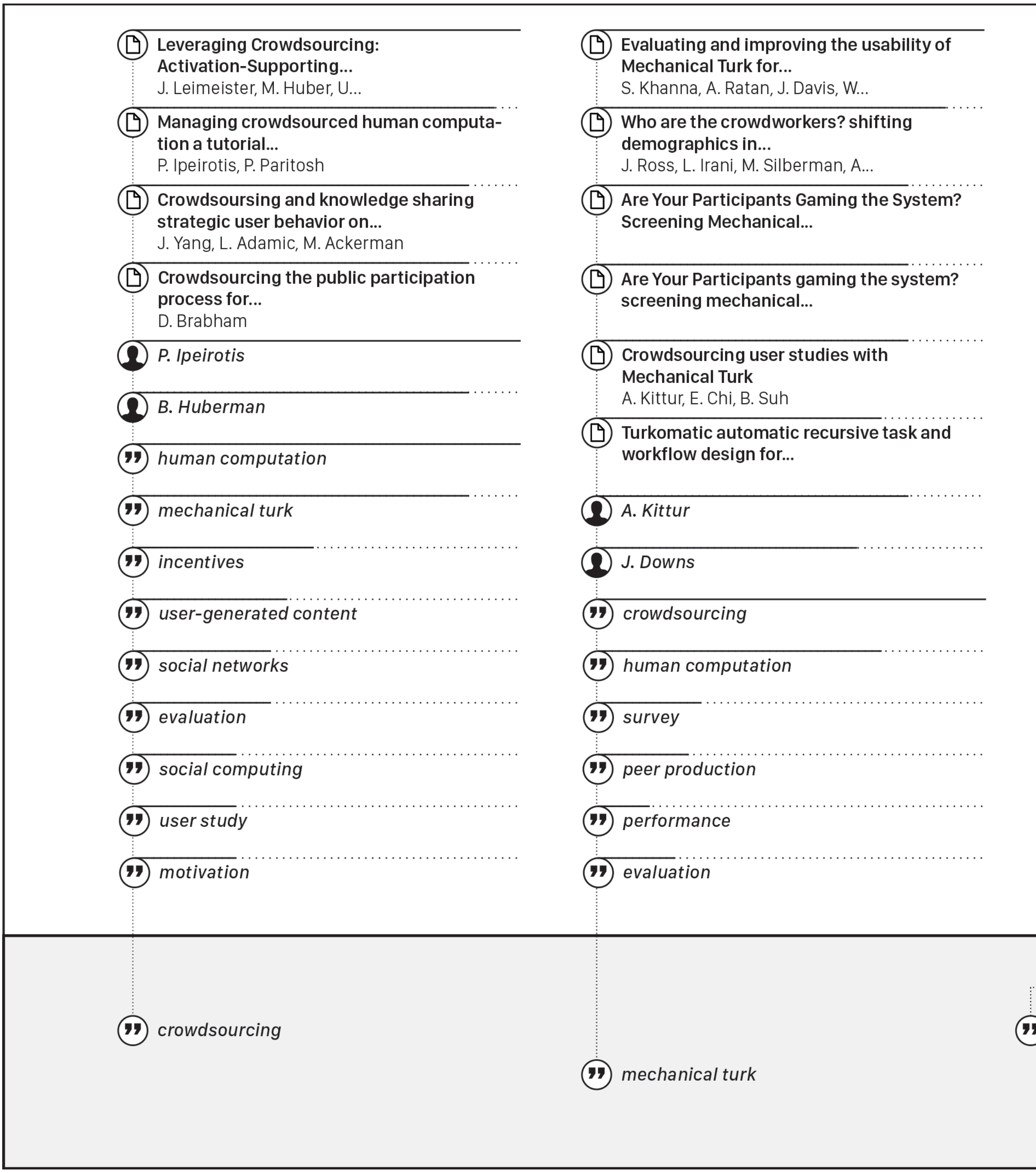
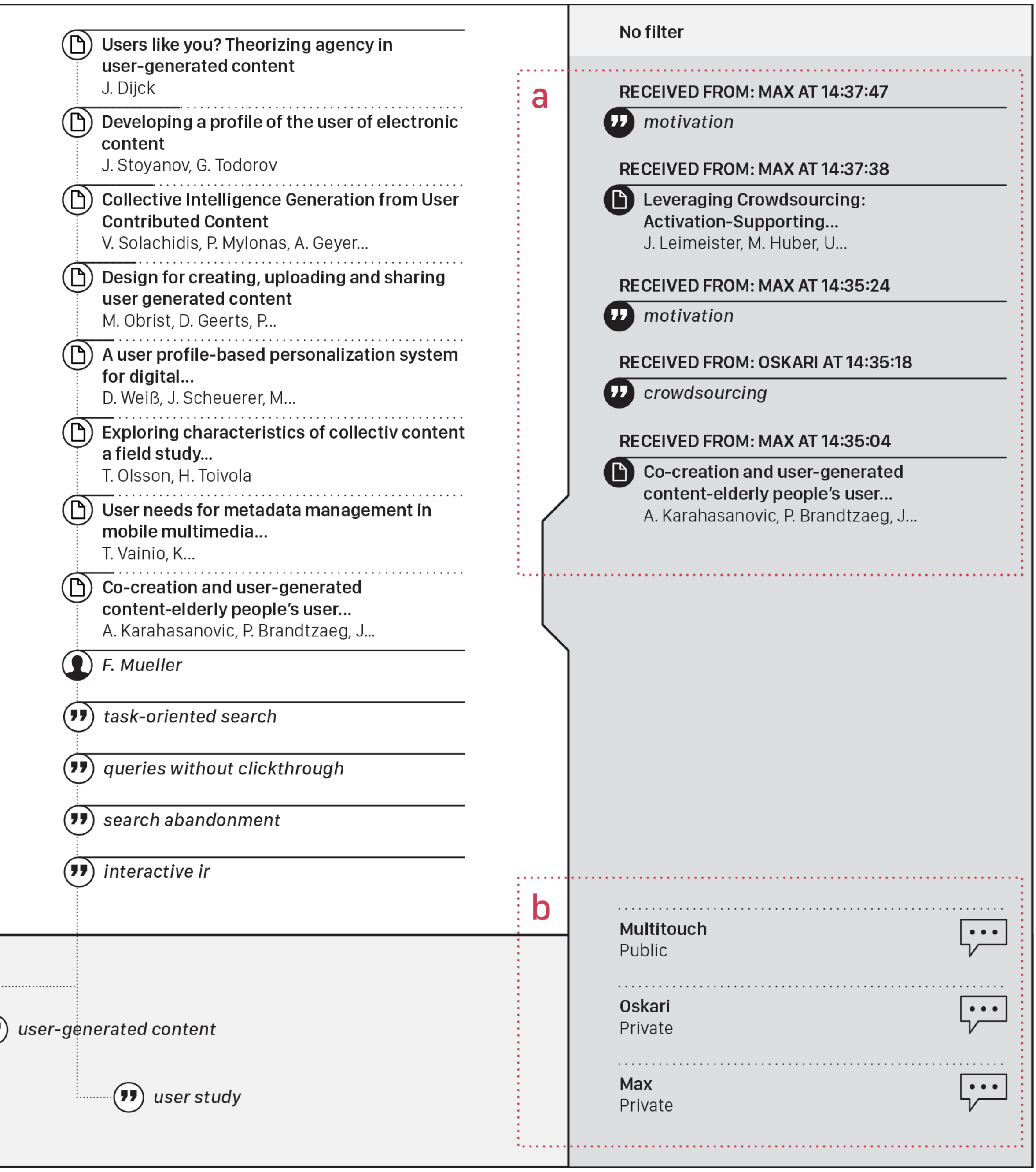


Figure 5.4. The user interface of QueryTogether. Differences with the ExplorationWall system include: (a) The reading list is now a share history and contains simultaneously entities bookmarked by the user, as well as entities received from collaborators. Their origin is indicated over the entity. (b) The user list shows online collaborators and the status of their device to differentiate at a glance between shared public screens and private devices. Entities can be shared by dragging them to the desired recipient in the user list.



Users like you? Theorizing agency in user-generated content
J. Dijck
Developing a profile of the user of electronic content
J. Stoyanov, G. Todorov
Collective Intelligence Generation from User Contributed Content
V. Solachidis, P. Mylonas, A. Geyer...
Design for creating, uploading and sharing user generated content
M. Obrist, D. Geerts, P...
A user profile-based personalization system for digital...
D. Weiß, J. Scheuerer, M...
Exploring characteristics of collectiv content a field study...
T. Olsson, H. Toivola
User needs for metadata management in mobile multimedia...
T. Vainio, K...
Co-creation and user-generated content-elderly people's user...
A. Karahasanovic, P. Brandtzaeg, J...
F. Mueller
task-oriented search
queries without clickthrough
search abandonment
interactive ir
user-generated content
user study
No filter
a
RECEIVED FROM: MAX AT 14:37:47
motivation
RECEIVED FROM: MAX AT 14:37:38
Leveraging Crowdsourcing: Activation-Supporting...
J. Leimeister, M. Huber, U...
RECEIVED FROM: MAX AT 14:35:24
motivation
RECEIVED FROM: OSKARI AT 14:35:18
crowdsourcing
RECEIVED FROM: MAX AT 14:35:04
Co-creation and user-generated content-elderly people's user...
A. Karahasanovic, P. Brandtzaeg, J...
b
Multitouch
Public
Oskari
Private
Max
Private

Sharing is performed by dragging an entity onto the chosen recipient in the user list, as seen in Figure 5.5. The recipient will instantly receive a new instance of the sent entity in her side panel. If the side panel is closed, a visual notification in the corner informs the user of the number of new entities received. Next to each user label, a “Message” icon allows the user to send a short message along with an entity.

To facilitate exploration based on saved and shared entities and documents, the reading list/share history can be filtered according to a chosen collaborator, simply by tapping his or her name. Filtering based on a collaborator that uses a personal device will show only entities and documents sent to and received from that user. Filtering based on the moderator, or any collaborator using a public workspace, e.g., through a shared monitor, large screen or projector, will display the content of the common collection shared among all users. Filtering based on one’s own name will display only entities that have been saved locally and ignore anything sent or received remotely.

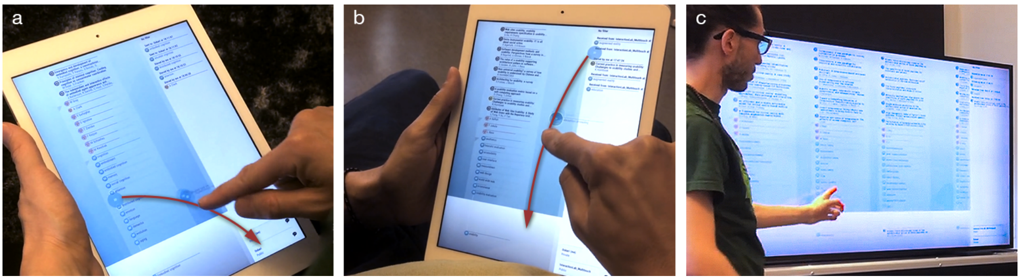


Figure 5.5. This sequence illustrates the coordination and exploration process between collaborators: a) From the results of a search, User 1 chooses an entity - in this case, a keyword - to send to User 2. User 1 drags the chosen entity to User 2’s label in the user panel. b) The received entity appears in User 2’s reading list along with information on the sender and the time it was sent. User 2 can then use it as an exploration trigger to start a new search stream by dragging the received entity to the query area. c) In the same way, entities can be sent to User 3. In this case, User 3 uses the system on a common workspace on a big screen and is set as public, so entities on the big screen are shared with all collaborators.

Example Scenario

Max, Anna, and Oscar are three computer science students. They have teamed up to present a common project in the context of a workshop on the semantic web. They meet to look for ideas as they only have a superficial knowledge of the topic. Max and Oscar take out their tablets, while Anna takes control of the shared large multi-touch screen, assuming the role of the public user and moderator of the session.

As they each log on to the collaborative search system, their names stack up in the user list on each device. They agree to start by exploring the topic of "semantic web" at large, to find inspiration and become more familiar with related subtopics. Anna starts by tapping somewhere in the query area of the shared screen, which opens a local text box. A soft keyboard pops up, with which she types "semantic web." Based on that query, the system returns a variety of related documents and keywords organized in a vertical stream. They discuss and review the keywords suggested by the search engine that seem to be the most central. They eventually agree that Oscar will further investigate the topic "ontologies" while Max will explore "web mining." On the shared screen, Anna drags the keyword "ontologies" from the result stream to the user label "Oscar" in the user list. A new instance of the keyword appears instantly in Oscar's side panel with the mention "Shared by Anna (Moderator) at 10:36:17." Anna then does the same with the keyword "web mining" to Max. Without having to type anything, Oscar privately drags the freshly received keyword on his device toward the query area, which returns a new stream of articles and keywords all related to "ontologies". He reads the abstracts of the retrieved articles and performs a few follow-up searches based on the related keywords. The new result sets appear as parallel streams on his interface, allowing him to compare their contents. In the same way, Max explores information related to "web mining."

As they both encounter interesting documents and keywords, they send them to the large screen by dragging them over to Anna's user label. After a little while, they decide to stop collecting new material to discuss the shared content. Anna leads the discussion on the large screen. As they review the outcome of their individual searches, they agree on which documents to keep in the list and which to dismiss. To remove an entity or a document from the list, they simply drag it out of the side panel. In the end, all three participants share the same collection of a few highly relevant documents that will make for an excellent basis to start their project.

5.2.3 Overview of the Study

We evaluated QueryTogether to investigate the type of improvement such a system could yield relative to a baseline condition that would replicate a realistic conventional collaborative exploratory setting. The study was aimed at providing understanding regarding the following aspects:

Figure 5.6. Experimental setup

Effectiveness: Does QueryTogether improve the search session effectiveness of retrieved information when compared to the baseline system?

Collaboration: Does QueryTogether enhance collaboration or lead to a more balanced contribution within the group of participants when compared to the baseline system?

Engagement: Does QueryTogether lead to more engaging search behavior when compared to the baseline system?

Additionally, we aimed to investigate potential changes regarding the use of searchable objects/entities, style, and heterogeneous devices:

Entities as searchable objects: What is the extent of use of entities as searchable objects in conversation or through the use of QueryTogether?

Collaboration styles: Does the collaboration differ in tightly or loosely coupled work styles or transitions between these?

Heterogeneous devices: Are there differences in how participants attend to the different devices?

The study followed a within-groups design with nine groups and two system conditions. Nine in-person teams of three people were assigned a collaborative search task (Figure 5.6). The system conditions were the full QueryTogether and a baseline version of the system that did not have the design features of QueryTogether. Each group performed two tasks:

one with the support of QueryTogether and one with the baseline. The conditions and tasks were counterbalanced by changing the order in which the two tasks were performed and the order in which the groups were subjected to each condition.

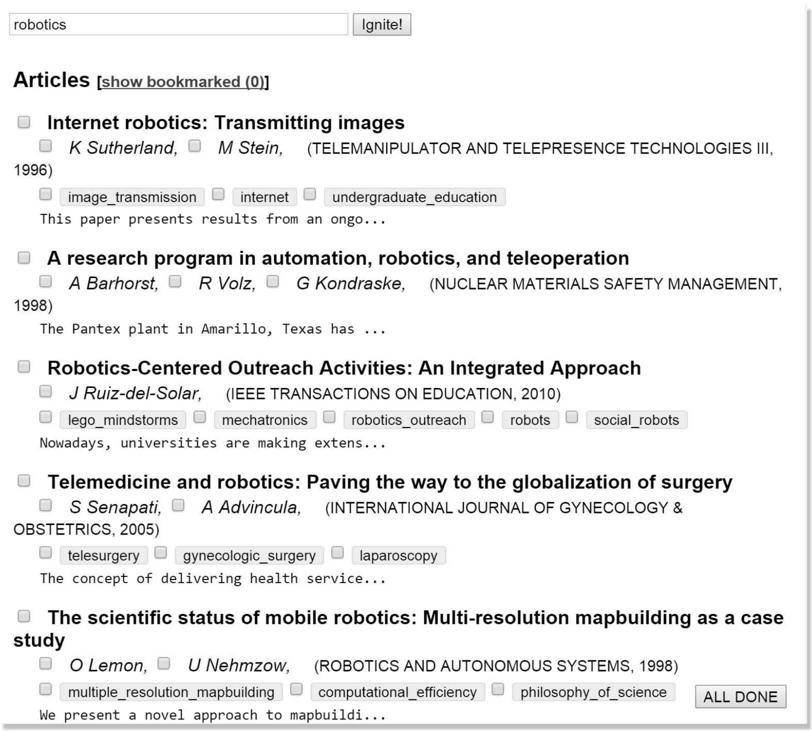


Figure 5.7. Baseline system replicating a conventional search interface while using the same backend as QueryTogether.

The baseline illustrated in Figure 5.7, was implemented based on this rationale. The baseline was a control condition that mimicked the conventional search interfaces, allowed for isolating confounding factors related to data or search engine functionality, and enabled the users' de-facto collaboration practice to be conveyed by combining the baseline system with the tools that the participants would use in real-life situations. We selected Google Docs as the information gathering and sharing platform, as this is the most commonly used collaborative platform available and the participants were likely to be familiar with the platform.

Two of the participants sat in chairs with an optional table, and the moderator sat at a separate desk. The three participants were placed in a triangle facing each other at a distance of approximately 2 meters to make it easy to see each other and communicate. When using the baseline, all participants used laptop computers. When using QueryTogether, two participants used tablets, and the moderator used a laptop computer. In both conditions, the moderator's computer was mirrored onto a large screen that faced the other two participants.

The participants had twenty minutes to complete the task with each system. After completing the task, participants were asked to fill out the User Engagement Scale questionnaire. The entire experimental session lasted about 80 minutes.

5.2.4 Findings

A first finding to highlight is a confirmation of the results obtained through the ExplorationWall study. As tasks pertaining to exploratory search are generally easier to perform on a desktop computer due to intrinsic limitations of touch-based devices [Singh et al., 2011]. Our measure of recall over time shows that QueryTogether, like ExplorationWall, allowed participants to be as effective with tablets as with laptops and conventional search tools. We also observed the same reduced need for typing, with 53% of search terms created by dragging and dropping entities into the query area, which emphasizes the utility of providing support for query formulation, and indicates that, when given the opportunity, people tend to prefer direct manipulation to typing on soft keyboards.

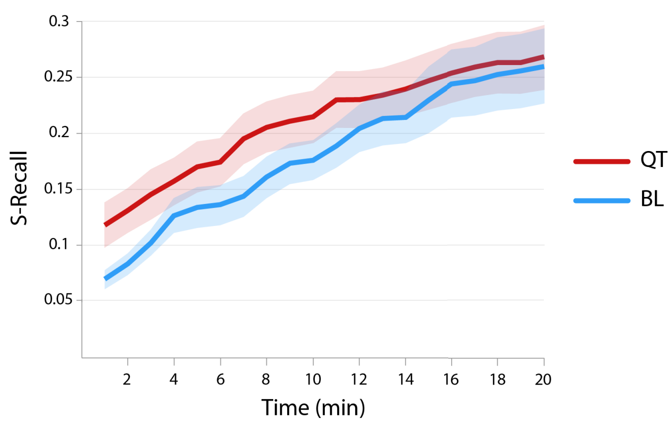


Figure 5.8. Measure of subtopic-Recall (S-Recall) [Zhai et al., 2015] over time averaged across groups. The curve shows how relevant elements were found earlier using QueryTogether, indicating a better distribution of work, with the positive consequence of leaving more time for discussions.

Analyses of effectiveness and verbal activity showed that most of the relevant subtopics were found sooner by participants using QueryTogether when compared to the baseline condition, as shown in Figure 5.8, which enabled more time for discussion and establishing a common ground. It also revealed how, with QueryTogether, collaboration strategies typical of collaborative exploratory search, e.g. division of labor or establishing a common ground, were better supported through the availability of entities as searchable objects. We also observed that QueryTogether supported a more flexible coupling [Dewan and Choudhary, 1991], meaning that the system made it easy to switch between phases of individual work with little coordination involved, and phases of tightly coupled collaboration, in which participants engaged in intense verbal communication.

Previous research shows *group awareness* as a positive factor in the coordination of actions, anticipation, and assistance provided to collaborators [Gutwin and Greenberg, 2002]. Our results suggest a higher group awareness has been achieved as the higher lookup rate at the shared display indicates. Moreover, the higher awareness of publicly shared information suggests the system supported the creation of a common understanding. Such interpretation is also supported by our results on verbal usage of entities, revealing that more time was spent in establishing common ground when using QueryTogether, by sharing prior knowledge on entities, asking clarification questions, explaining, and reporting activities performed with entities.

5.2.5 Design Implications

The main question regarding the design of QueryTogether was the use of entities to demarcate information spaces collaboratively. **Participants successfully shared entities as a means to understand each other and find common-ground**, they also were able to naturally come up with flexible strategies to alternate between individual and collaborative work, on top of confirming individual benefits observed using ExplorationWall in section 4.1.

5.3 Contribution

The aim of this chapter was to answer RQ2: *How to demarcate and visualize a coherent information space through entity-based affordances?*. User experiments with RelevanceMap showed how mapping of information can be achieved to provide users with the means to perceive the information space in a way that is semantically consistent. As a result, we observed participants performing effective orientation leading to efficient discovery of information related to complex criteria. User experiments with Query-Together allowed us to better understand how entity-based affordances could support exploration by multiple users collaborating in a common information space, helping them distribute the work and find common ground. Together, these results provided satisfactory answers to RQ2.

6. Enabling Entity Affordances Across Environments

We have established that information exploration is long-term, and relies on insights from many sources, including active search, serendipitous finds online and offline, conversations, etc. However, the lack of consistent and direct communication between various sources of information results in a burden for the user who must overly rely on her memory to make sense of encountered information and get a sense of context out of it, which creates few opportunities for insight and contributes to unnecessary cognitive load. We describe this problem as pertaining to continuity in the information space.

It is clear how the affordances proposed so far can function within one application. However, to be useful, such affordances should potentially be harnessed beyond such a limited environment. For that reason, the next logical step is to investigate the potential for enabling these techniques across various applications, e.g., browser, email client, ebook reader, but also beyond the digital realm, by exploring techniques to channel physical or temporal resources, such as conversation-sourced information, into entities to be interacted with.

We decided to address this continuity problem through the implementation of proactive systems, able to monitor an active situation, e.g., a conversation, writing of a report or planning of a trip, and from the data acquired, the system proposes entities that can then be interacted with, shared, or integrated to a visualization and used to steer and further the exploration.

This chapter reports on three separate systems. The first two, summarized in Sections 6.1 and 6.2 and reported in Publication IV and Publication VI, use speech recognition technologies to provide proactive help and entity recommendation to users engaged in natural conversations. The last system, summarized in Section 6.3 and reported in Publication VII, visually monitors a user's activity across digital applications, makes inferences regarding the tasks at hand, and finally proactively provides recommendation on various entity types. The three systems are summarized in the following sections, as well as their respective evaluation. Together they provide a rich exploration of techniques for enabling continuity in the information space, therefore providing satisfying answers to RQ3: *How to benefit from entity-based interactions for exploration beyond self-contained systems?*.

6.1 InspirationWall: Topic Recommendation Through Speech Recognition

As a proof of concept, we first wanted to develop a minimal system that would positively influence the outcome of a conversation, as a basis for more complex systems. With that in mind, we settled on augmenting an ideation – or brainstorming – session. In such sessions, participants try for a limited time period to come up with as many ideas as they can on a given topic or with a few given constrains, while deferring judgment and avoiding censoring themselves. Such focus on quantity over quality often yields silly ideas that can however trigger insights and novel associations. The same focus on quantity over quality is also something that can easily be emulated by a digital agent, that would use expressions pronounced by human participants as queries to retrieve additional topics that would then fuel the session. As unexpectedly interrupting the participants would obviously be counterproductive, we strived to design a system that could be consulted any time with minimal effort, while being completely unobtrusive when not needed. We were interested in understanding if, in such a limited setting, participants would make use of the system, and if input from the system would influence the session in any positive way.

6.1.1 Overview of the System

InspirationWall has been designed to be a supplementary participant in group ideation meetings. A system that "listens" and contributes in a way that does not interrupt the flow of ideas. Therefore, the system required minimal interaction from users/co-participants. The visible part of the apparatus consists of a screen and a microphone. The system continuously monitors an ongoing conversation. The audio is transcribed into text in real

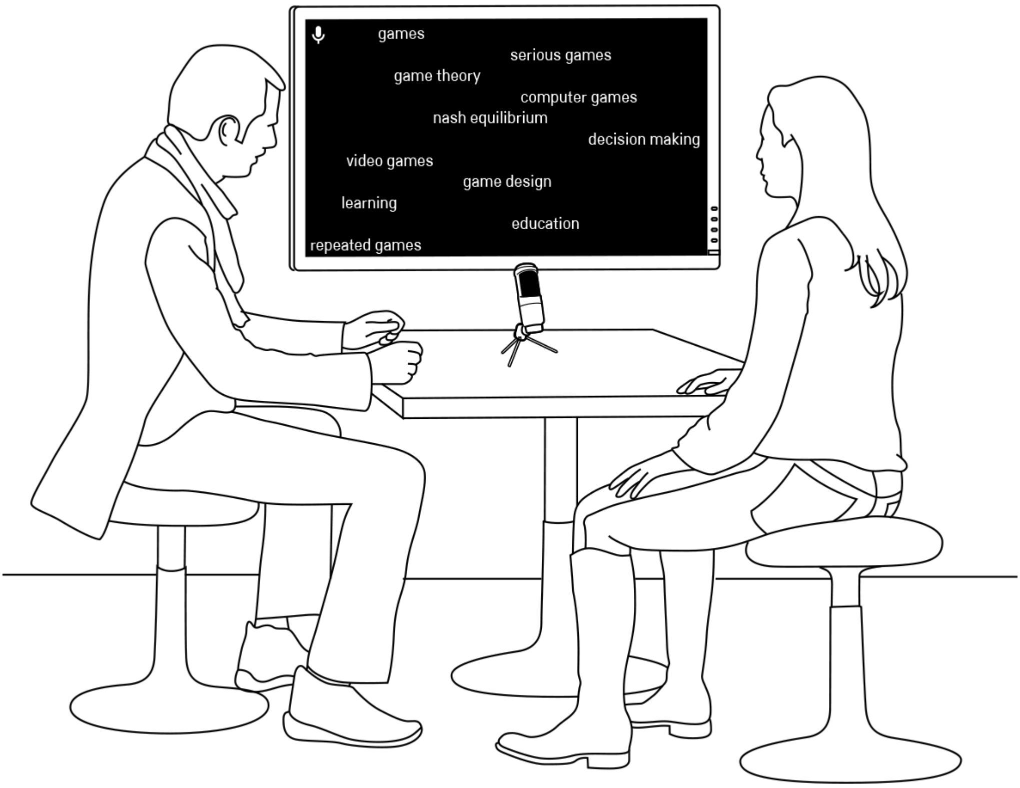


Figure 6.1. User Interface *in situ*. Words related to the ongoing conversation fall slowly across the display, ensuring a constant flow of new topics.

time[1]. From the text, significant words are isolated and used as a query in a search system – which uses the same backend as ExplorationWall – to retrieve related keywords to be displayed.

Visually, the InspirationWall (Figure 6.1) user interface has a dark background with, at first, nothing but a small microphone icon blinking in the upper left corner to indicate the system is active and listening. When significant words are heard by the system, and related keywords found, these are displayed in a silent "rain of words": words appear at fixed intervals from the top edge of the screen, at a random horizontal position, and moves down at a slow, fixed speed. This provides a slow continuous stream of words to potentially enrich the ideation session, without disturbing it through sudden changes or sounds. The system is conceived as an aid, that

[1]The technique used for automatic transcription of natural speech is detailed in Publication V [Andolina et al., 2015a].

Figure 6.2. Experimental setup

can be consulted or ignored depending on each participant's need.

6.1.2 Overview of the Study

We conducted a study with six pairs of participants to evaluate the effects of InspirationWall on the idea generation process, with the goal of understanding whether it helped small groups generate more ideas. Each pair took place at a table, with the display on one side, as if it was a third participant (Figure 6.2). Paper and pens were provided. They had twice twenty minutes to come up with as many project ideas as possible on a given topic. Once with InspirationWall being active, and another time without the help of the system. We used two different topics, i.e., *robotics* and *wearable computing*, and we fully counterbalanced both the order and topics used for both conditions. Participants were recorded through the same microphone used by the system for the speech-recognition, and a camera at the top of the screen was taping each session. The alignment of the camera with the display made it easy to identify occurrences of participants glancing at the screen.

We measured *quantity of ideas over time*. Using video and data logs, we were able to identify individual ideas and the approximate time of inception. We also counted the occurrences of participants looking at the display to identify when an idea could be linked to something that was on display at the time.

6.1.3 Findings

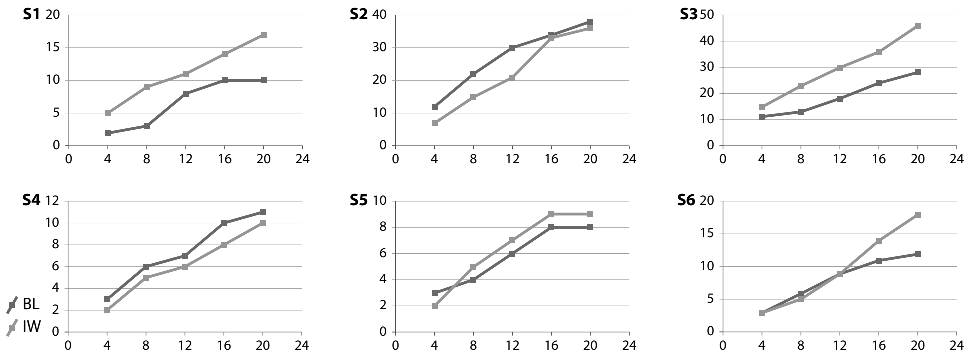


Figure 6.3. Accumulation of ideas per condition (BL = Baseline; IW = InspirationWall) in the different sessions S1,...,S6. On the Y-axis is the cumulative number of ideas, and on the X-axis is the time from the beginning of the session (minutes).

Participants were asked to generate ideas but not explicitly to use or interact with the system which was simply provided as-is. Our study shows that participants that used InspirationWall more – as indicated by the count and duration of gazing occurrences obtained through video analysis – tended to generate more ideas in total and over time, as seen in Figure 6.3. Those results suggest that InspirationWall contrasts the decay of idea productivity over time – typical of traditional idea generation sessions – and confirm the effectiveness of automatic information exploration and keyword suggestion on idea generation.

6.1.4 Design Implications

InspirationWall is constitutes a preliminary, quick and dirty implementation of a proactive system monitoring conversations. However, our study shows it as **an example of an entity-oriented system that is able to provide support in information practices taking place outside a digital environment**, which opens a variety of directions for future work, including allowing richer interactions with such systems.

6.2 SearchBot: Proactive Entity Search

Following the encouraging results of InspirationWall, we were interested in further exploring how a digital agent could support/influence a conversation by providing proactive entity recommendations and contextual information. To that end, we needed the setting to be more common and naturalistic than an ideation session, and we wanted the system to allow more interaction with the presented information. Therefore we settled on a natural conversation setting and aimed to develop a system that would listen to the conversation and provide individual understanding support by showing entities and links related to what is being discussed, enabling quick contextual on-demand information.

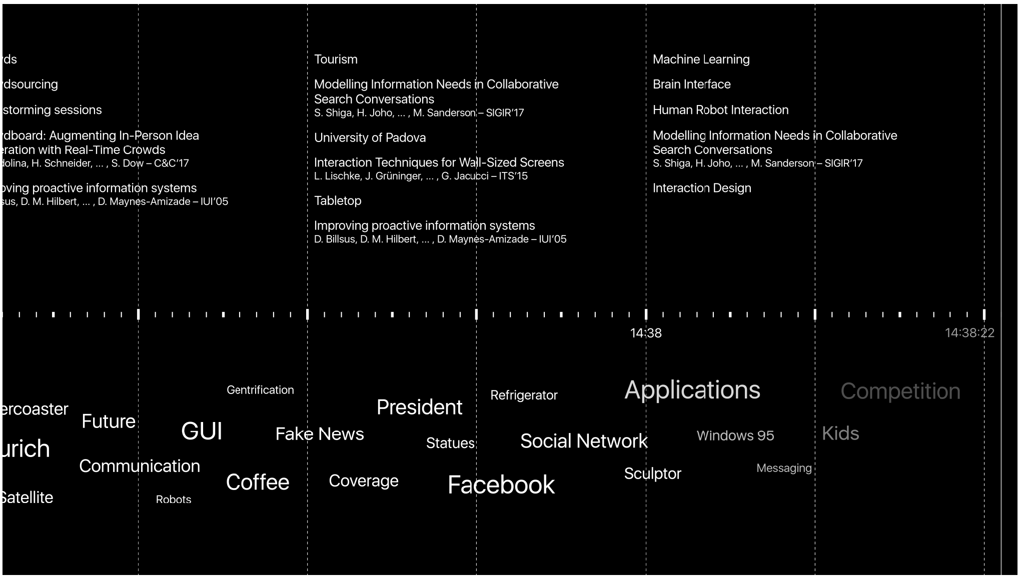
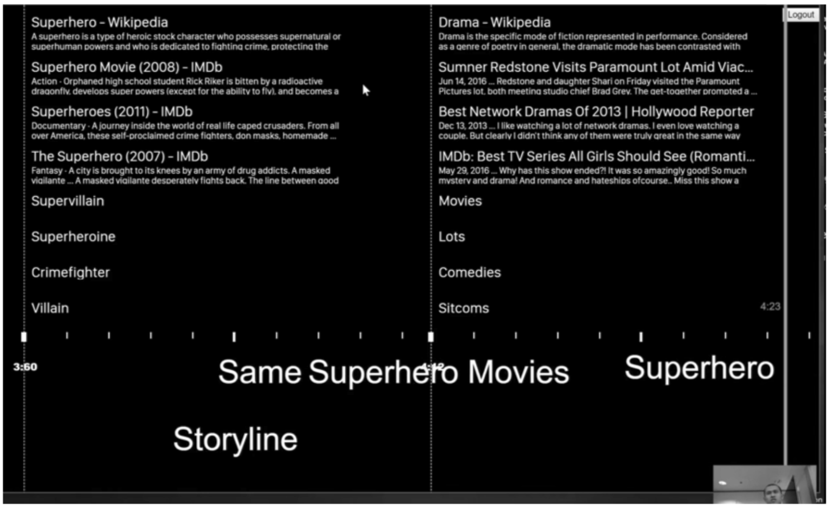


Figure 6.4. The user interface of the SearchBot system. The system monitors a conversation and provides continuous recommendations of related documents and entities in a non-intrusive way: On the lower half, a continuous stream of recognized entities; In the middle, a timescale with timecodes; On the upper half, sets of recommended entities, i.e., documents and keywords. The user can go back by sliding the timescale to the left to retrieve past recommendations, and return to *the present* by sliding again to the right.

6.2.1 Overview of the System

We designed the SearchBot system to monitor a conversation and provide continuous recommendations of related documents and entities. SearchBot listens to conversations through a microphone and a speech recognizer (i.e., Google Web Speech API). Each detected sentence is then processed by Google's Cloud Natural Language API, which detects entities in the sentence and uses them for the recommendation process. Related documents are retrieved via Google Custom Search with detected entities as a query.



…
Participant 1: Yeah. For superhero movies I was very into it because I just-- you know when I was little boy they were all, they were all around me so…
Participant 1: That was-- that was when I thought, "Wow, a movie can be made like this".
Participant 1: Oh that's cool. The-- besides super hero movies, I'm also into dramas.
…

Figure 6.5. Example screen capture of the system in a session where the participants were having a natural conversation about movies. The corresponding transcripts of the spoken conversations are shown below the screen captures. The system is recognizing and recommending entities and matching documents based on the conversational input.

An example of a sentence, extracted entities, their types, and the resulting queries:

```
Sentence:  Bordeaux is famous for its wines.
Entities:  Bordeaux (type location), wines (type consumer good)
Named entity query vector:  Bordeaux
General query vector:  Bordeaux + wines
```

The system runs in a regular browser window (Figure 6.4). It consists of a timeline that displays a stream of recognized entities in the lower part of the window, a timescale with timecodes displayed in the center, and successive sets of four retrieved documents and four recommended topics in the upper part of the window. A new set extends the timeline every time a new transcription is available.

The user can interact with the system in multiple ways. Clicking on recognized or recommended entities triggers a search and opens the most relevant article in a new tab. Clicking on a document will open its content

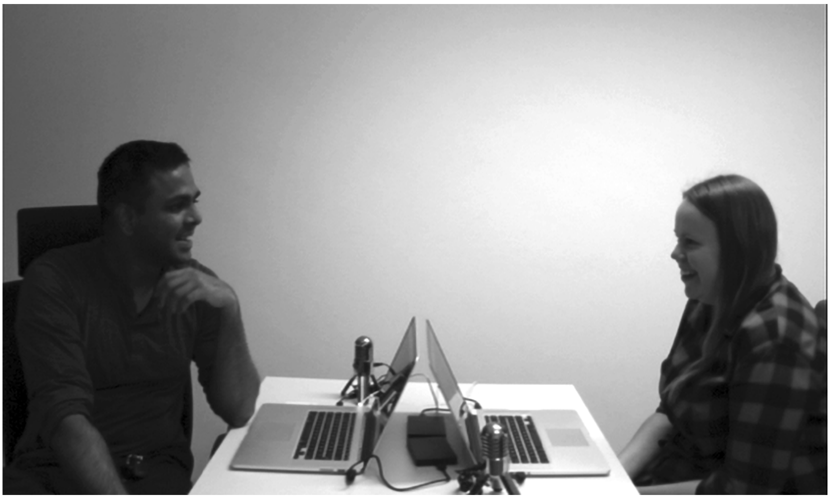


Figure 6.6. Experimental setup. Participants were sitting around a table, and a laptop was placed in front of each participant. The laptops were displaying the SearchBot interface. Microphones were placed on the table to record the conversation.

in a new tab. Users can also move back and forth in the timeline by clicking on and dragging the central portion of the window.

An example of a system screen captures during a spoken conversation in which participants were having a natural conversation about movies is shown in Figure 6.5.

6.2.2 Overview of the Study

We aimed to investigate how a proactive search agent can support natural spoken conversations between people by augmenting the conversations with additional information. To that end, we conducted a user experiment with 12 pairs of participants in a within-subjects design. Each pair of participants took part in twice twenty minutes of informal conversation, once with access to SearchBot and the second time without SearchBot but with access to a custom version of Google Search to avoid any confounding variable regarding variation in the quality of retrieved information. Conversations were kick-started through a suggested topic, i.e., *Movies* or *Travels*. Both topics were fully counterbalanced with respect to the condition in use, and order of sessions.

In order to assess the SearchBot's effect on the conversation, we used objective and subjective measures.

Influence of information shown on the conversation. To understand whether the information presented on the screen influenced the conversa-

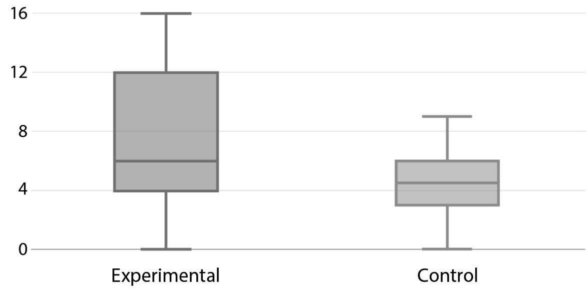


Figure 6.7. Number of entities that were mentioned within 60 seconds of when they were shown on the screen.

tion, we counted the entities extracted from the items shown on the screen that were mentioned in the 60 seconds following their first appearance on the screen. To control for possible cases in which displayed entities were mentioned by chance, we performed the same calculation in the control condition. In this case, the proactive search interface was running in the background, and results were not shown to the participants.

Consumption of web resources. The number of pages opened by participants during the conversation served as a proxy for the consumption of Web resources.

Perceived quality of the recommended items. We showed participants the list of the 100 recommended entities and the list of the 100 Web documents that the system displayed most frequently, and we asked them to mark the items that they considered pertinent and relevant to the conversation. We considered this measure a proxy for the perceived quality of the items the system suggested.

Preferred items with the proactive search agent. In the experimental condition, we logged the item types (i.e., Web documents, recommended entities, and recognized entities) that the system displayed and that the user clicked on to seek more information.

Subjective experience. We investigated participants' subjective experiences with the system using a questionnaire and a semi-structured interview.

6.2.3 Findings

Results show that participants in the experimental condition frequently referred to the entities and documents shown on the screen during their conversations (Figure 6.7). The comparison with the control condition, in particular, demonstrated that these references were not due to chance.

This result indicates that not only did the proactive search system retrieve useful information, but the displayed information influenced the conversation, as questionnaires and interviews further confirmed.

There was no significant difference in the number of Web resources consulted between the experimental and control systems. This result suggests that participants retrieved the same number of useful resources supporting the conversation in both experimental conditions. However, while in one case the resources were automatically retrieved by the proactive search agent, in the other case, explicit query formulation and refinement was needed.

In general, the reported quality of the experience of using the system was more positive for the experimental condition, as it allowed participants to keep eye contact with each other, enabling more fluent conversation. Participants reported that SearchBot allowed them to check facts and build common ground without needing to exert much mental effort. Furthermore, the system was able to expand the conversation in new directions. However, the added value of the proactive search experience seemed to come with the cost of feeling less in control of the search process.

6.2.4 Design Implications

These results show that entity-based interaction can be engaged successfully in conversations, as a way to support an interlocutor's immediate understanding, or as an advanced method for automatic note-taking. Future work could investigate, how **a proactive system could present, after a conversation has taken place, a summary of it in the form of a network of entities**. Such visualization would provide an overview of the different topics, opinions, entities of interest, in other words, not only a transcript but the context of the conversation displayed as interactive entities ready to be used in queries, organized or shared.

6.3 Entity Recommendation Across Digital Environments

So far, the systems described in this chapter have presented opportunities for harnessing entities outside of digital environments through conversation monitoring. An important remaining challenge consists of enabling entity affordances across multiple digital environments, allowing a user to search, read, write, communicate using a variety of digital tools while still taking advantage of entities.

6.3.1 Overview of the System

User Interface

The system's user interface is illustrated in Figure 6.8. It implements three specific features:

1. Showing entities being recommended by the system

2. Allowing selection of entities of interest by the user (explicit feedback)

3. Allowing direct action on entities when relevant

In the following, we describe how each of these features were implemented in our experimental setup. Finally, we provide a use case example demonstrating the use of the entire system.

Recommended entities are displayed within four rows of five items, one row per entity type, i.e., people, applications, documents and topics (as keywords).

People are identified by their name under a photo-based icon when available, and a standard anonymous silhouette when not. Applications are identified by their names under a standard icon or logo of the application or service. Documents are identified by their name under an icon based on a preview of their content, with a small icon of the application used to read or edit it. Finally, topics are identified as a single keyword. In each row, recommended entities are ranked horizontally from left to right. Since the main purpose is to show a small variety of the most relevant entities, the ranking is not visually emphasized. As users perform their tasks, the system progressively updates the recommendations. These changes are reflected in the UI as entities eventually shift places and new entities replace old ones in each row. In the prototyping phase, since entities are displayed on an orthogonal grid, some users tried to derive meaning from the vertical alignment of entities across rows. To prevent that, the grouping of recommended entities by type in each row has been emphasized with a grey rectangle that acts as a container.

When the user is interested in a specific entity among the recommendations, she must be able to express her interest in a way that informs the

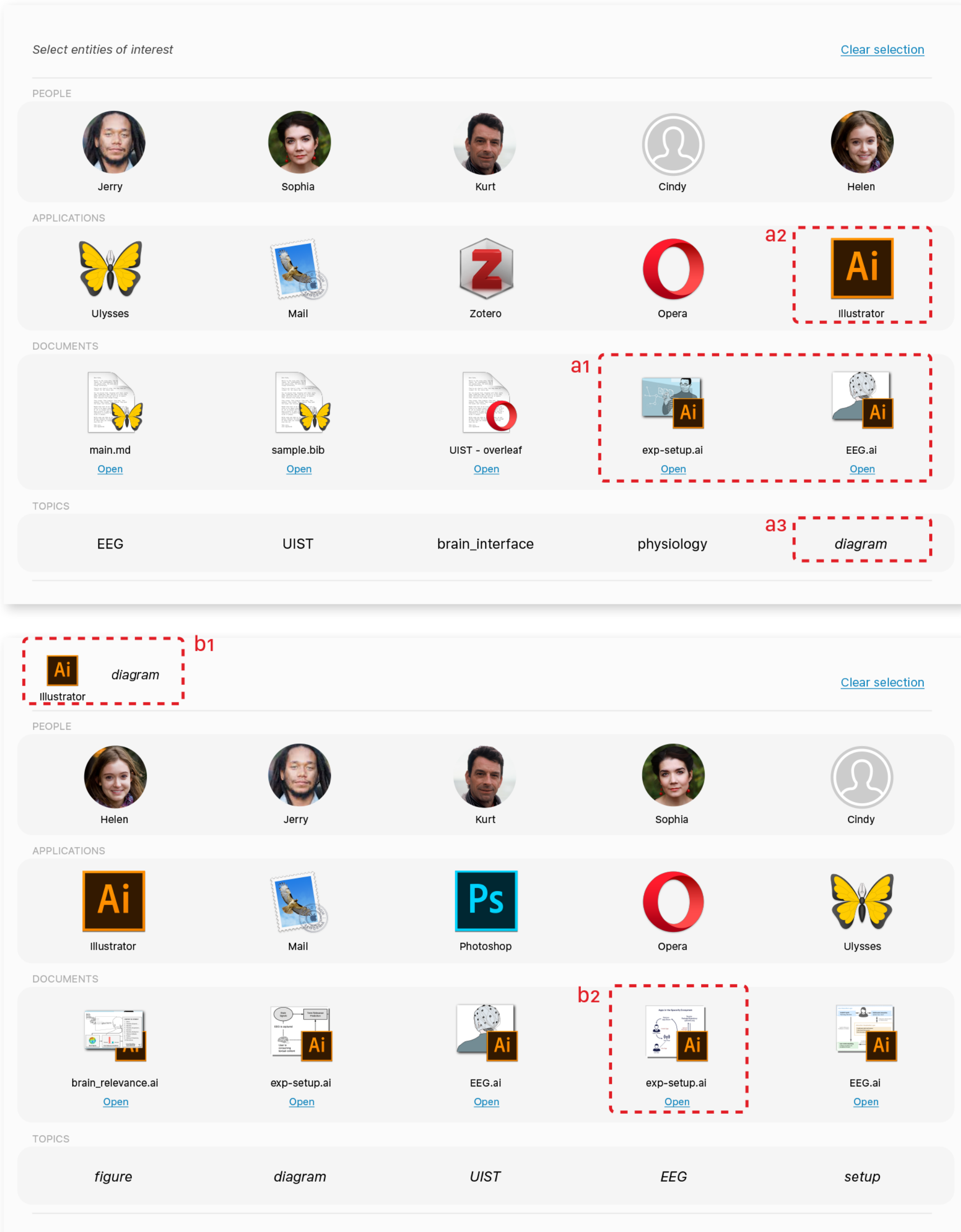


Figure 6.8. Two states of the system's user interface. Recommended entities are displayed within four rows, here with five items each: people, applications, documents, and topics. The user can select entities of interest by clicking on them, which updates the recommendations. Up: the user sees entities related to her current work. She notices figures she has made for one of her papers (a1). She clicks on "Illustrator" (an application for editing vector graphics) (a2), then on the topic "diagram" (a3). Down: As a result, the entities of interest are displayed in the top area (b1) and the system updates the recommendations accordingly with the user's selection. In the document row, she selects an illustration (b2) that she will modify for use in her new paper.

system so that recommendations update accordingly. To that end, every recommended entity displayed on the UI can be selected with a click. As a result, the selected entity, or entity of interest, appears in the area at the top and the overall recommendations (in every row) are updated, taking the selection into account (i.e., a positive feedback on the selected entity is sent to the system). More entities can then be selected and added to the entities of interest at the top of the screen, providing an explicit way to influence the recommendations. Entities of interest can be removed from the selection by clicking the cross that appears at their upper right corner when the mouse cursor hovers their icon. Removal of an entity of interest from the selection sends a neutral feedback on the selected entity to the system, which updates the recommendations accordingly. The whole selection of entities of interests can be reset by clicking the "Clear selection" button on the right.

An important feature of the system is to make the recommendations actionable. While work on translating recommended persons and keywords into potential actions is ongoing, the present version simply allows to directly open recommended applications and documents. Figure 6.9 illustrates the user interface through an example scenario.

Example Scenario

Alice is evaluating a Master's Thesis about interactions in virtual reality. The work is quite interesting and she's almost done with the first pass of annotations. A notification pops up in the corner of her display: it is an email from the university administration stating the budget she has submitted for a conference trip next month has been approved. This makes her think she better hurry if she wants to be able to choose which flights and hotel will better suit her needs. Prioritizing the new task, she interrupts the evaluation and opens the travel portal. After a few seconds, the entity recommendation display gets updated with entities of the budget document she sent last week, the concerned administration person, as well her colleague and co-author Bob, who is supposed to give a talk at the conference as well. By glancing at the screen Alice sees the name of Bob and she realizes he could provide useful advice for what concerns which hotel to stay as he had already planned his trip. She opens a new direct messaging conversation. "Hi there. Do you already know in which hotel you'll be staying in Hong Kong?". Bob answers and shares some points of interest he plans to visit. With that information, Alice finishes her

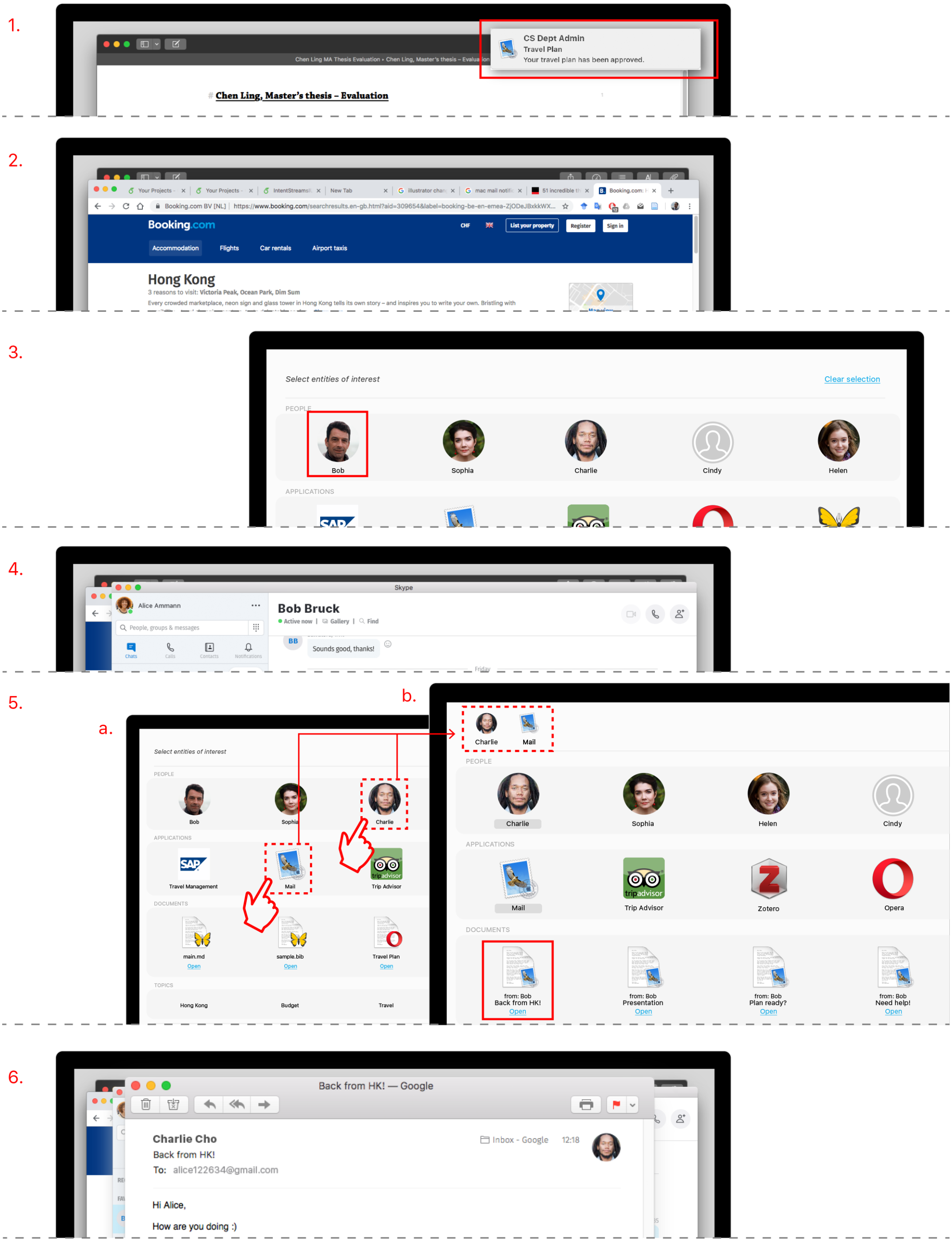


Figure 6.9. Use case scenario.

booking. Now intrigued, she resumes her exploration of what to do in the city when she's there. The entity recommendation screen now shows related trip advisor links, but the system also recommends her friend Charlie. That recommendation makes her recall Charlie had once sent an email describing his own trip to Hong Kong. Looking for that email, she selects Charlie and the email client, which tells the system to focus on items related to these entities in the following recommendations. With the provided feedback the system manages to dig up Charlie's old emails, which included some personal recommendations that have become useful at last. Alice starts to be pretty excited about the trip but realizes she has an urgent matter to attend. She clears the selected entities and resumes her thesis evaluation task. After a few seconds, the system shows again an entity-based overview of her current tasks, including the submitted Master's Thesis manuscript and her annotated document.

6.3.2 Overview of the Study

A user experiment was conducted to assess the quality of recommendations and understand how it influences users' behavior and subjective experience of the task at hand. The study followed a within-subjects design, with two conditions, one with the recommender system visible, and a control condition with the same system running in the background without being visible to the participant. Thirteen participants took part in a two-phase experiment: two-week digital activity monitoring, in which participants had to keep a diary of all their digital activities, while the system's logging software was running in the background on their personal laptop, followed by a controlled lab study (see setup on Figure 6.10) in which participants had to perform two tasks picked from their diary, e.g., *course preparation*, *literature review*, *programming* or *travel planning*, using each system condition for ten minutes. The order of both conditions mas fully counterbalanced amongst participants. For the lab experiment, participants' laptops were hooked to a supplemental display, which showed the entity recommendations in one of the conditions, and to a calibrated eye tracker that would record gaze data. After the task, participants were interviewed to register their subjective experience of relevance and influence of the system, then had to assess the relevance of entities that were recommended to them during the session. The influence was measured through duration of gaze fixation on recommended entities, explicit interaction with recommended entities and direct utilization of recommended information.

Figure 6.10. The interactive setup. Participants used their own laptops to perform the tasks. An external monitor was set up to connect to the laptops showing the recommender system's UI. SMI eye tracking device was installed and mounted onto the external monitor to track participants' eye gaze behavior during the tasks. In the figure, a participant continues a writing task for a research paper while the recommender system continuously suggests relevant references to the manuscript.

6.3.3 Findings

Our results showed that the proactive recommender system effectively supported users in performing their tasks. The system was able to accurately extract context across applications. The proposed entity interactions, with the added context leverage, often helped participants retrieve useful items for which they could not recall a specific pointer, e.g., title, source or location. The simple possibility to select an entity of interest to orient the recommendations proved effective, with data showing that such affordance was consistently used during the session, and our qualitative findings suggest such affordance was an important factor in the overall positive experience using the system.

6.3.4 Design Implications

Beyond immediate use, as the entity recommendation was used here, this prototype and study show how a person's activity can be followed in the background by a proactive agent, which would transcribe these activities into an entity-oriented log, **allowing the user to fluidly transition into an entity-oriented environment, utilizing her recent insights**

as material to enrich ongoing or past explorations.

6.4 Contribution

This chapter aimed at answering RQ3: *How to benefit from entity-based interactions for exploration beyond self-contained systems?*. InspirationWall demonstrated the possibility to influence the outcome of collaborative ideation sessions using speech recognition to leverage context, implemented into a non-intrusive setup. SearchBot also used speech recognition in a conversational setting with the added possibility to interact with recommended entities to access more contextual information and help find common ground with the interlocutor. Finally, we demonstrated the utility of a system proactively recommending entities in relation to a user's task at hand.

It must be noted that proposing systems that infer information of interest based on implicit inputs would not do much in terms of improving user control over encountered information. Transparency-wise, recommender systems represent a less than ideal form of support for information seeking. Our use of proactive agents in the tested systems allow us to evaluate the effects of the possibility to incorporate traces of relevant information extracted from temporal activities, e.g. conversation or everyday tasks, into the greater exploration activity, in the form of actionable entities that can be actively shared, organized and used into queries. The dangers of automated recommendation could ideally be avoided through the comprehensive extraction of entities in such settings, which would yield a large amount of information to be visualized and explored using the same techniques described in chapters 4 and 5.

Instead, the main contribution of such systems lies in the possibility to incorporate traces of relevant information extracted from temporal activities, e.g. conversation or everyday tasks, into the greater exploration activity, in the form of actionable entities that can be actively shared, organized and used into queries. As such, the three described systems and respective studies have demonstrated that entity affordances proposed in chapters 4 and 5 do not constrain a user's exploration within a single dedicated digital application and can be used as a way to augment various aspects of it and to centralize respective outcomes, thus addressing RQ3.

7. Discussion

In this chapter, we reflect on our exploration of entity-based interaction techniques designed to support information exploration. First, I summarize the main findings of the presented research. Then, I propose a design template describing the Hypercue [Klouche et al., 2018], an interactive representation of entities that provides personalized access points to information and serves as a complement to hyperlinks. The Hypercue design template consists of a minimal set of affordances that ensure all important features for supporting exploratory search can be addressed while leaving enough design space to facilitate its integration in a variety of systems. Finally, we discuss the implications of our work and propose future work directions.

7.1 Summary of the Main Findings

The work presented in this dissertation aimed to investigate the design space of entity-based explorability of the information space, following three properties: *Direction*, *Orientation*, and *Continuity*.

Chapter 4, which summarized the work reported in publications I and II, aimed to address *Direction* by answering RQ1: *How can entity-based querying benefit information exploration?* Both systems described, ExplorationWall and IntentStreams, successfully demonstrated techniques for directing exploratory search through entity affordances. Study results showed, in the case of ExplorationWall, that the proposed solution, dubbed *Entity Picking*, had positive effects on search activity, behavior, and engagement in exploratory search tasks, and even outperformed conventional text-based querying in search performance. IntentStreams introduced support from an intelligent agent in the form of *interactive user intent modeling* that coupled relevance feedback with visualization of the user model to keep the user in control. The significantly positive results obtained regarding both performance and engagement confirmed that the proposed interaction techniques were effective for directing exploration, therefore addressing RQ1.

The work summarized in Chapter 5, reported in publications III and IV, aimed to address *Orientation* by answering RQ2: *How to demarcate and visualize a coherent information space through entity-based affordances?* First, RelevanceMap showed an example of a user-driven visualization of an information space that helps users more accurately demarcate and perceive an information space using situated entities on a 2D map. Second, QueryTogether showed how users took advantage of entity-based affordances in co-located collaborative exploratory tasks using multiple devices, to coordinate their team and find common ground. These two complementary cases showed how entities can be used interactively to demarcate a coherent information space of interest, thus answering RQ2.

Chapter 6, which summarized the work reported in publications V, VI and VII, aimed to address *Continuity* by answering RQ3: *How to benefit from entity-based interactions for exploration beyond self-contained systems?* InspirationWall and SearchBot showed two examples of background conversation monitoring for proactive entity recommendation. SearchBot in particular demonstrated the feasibility of real-time relevant recommendations from such conversation monitoring. The last presented system used another type of monitoring, i.e., digital activity analysis, to infer the task in which the user was engaged and provide relevant entity recommendation and possible related actions. Beyond the setting of these cases, the successful monitoring of a user's activity outside the boundaries of one dedicated digital application demonstrated the potential for entities to be harnessed in various contexts pertaining to the complex process of information exploration, thus providing a satisfying answer to RQ3.

7.2 From Hyperlinks to Hypercues: Integrating Entity Affordances

Following the research findings, I reflect in this section on my exploration of interaction techniques designed to support entity-oriented information exploration. This reflection is summarized from Publication VIII. That last work aimed at devising a minimal set of affordances that ensure all important features for supporting exploratory search are addressed while facilitating their integration in various existing applications. A driver for such reflection, and a followup to my three research questions, was the following new question: *What would be an equivalent to the Hyperlink in the proposed paradigm of entity-oriented exploration?*

My answer is dubbed the *Hypercue*, an interactive representation of entities that provides personalized access points to information and serves as a complement to hyperlinks. Hypercues create opportunities to flexibly discover, store, and share information, and to gain insights from the data. We describe the rationale behind the design template for Hypercues and discuss its implications.

A cue is a stimulus and a signal for action. A hypercue is an interactive representation of an entity; it offers affordances for the user to explore, share and organize her thoughts. Systematic inspection and exploration of the design space of each feature of exploratory search systems allowed us to identify three complementary affordances that are responsible for enabling these features and that together constitute a minimal design template for implementing hypercues. The following template aims to guide the creation of future interfaces for exploration without overconstraining the design of such systems, or hindering the ability to address specific cases through the choice of a specific form of visualization. The proposed affordances can also be implemented in most existing media-handling applications (e.g., in browsers, PDF and e-book readers). From the user's perspective, the following template provides a base set of rules and expectations to facilitate users' engagement in complex information behavior.

Affordance 1: Entity-Based Querying

Each entity or combination of entities yields various new related entities, thus providing an overview of the respective information space.

Providing the ability to create queries through the direct manipulation of recommended entities can **support query formulation and facilitate query refinement**. The ability to add more entities to an initial query makes it possible to refine it by narrowing down or expanding the result set. Adding external entities (e.g., from somewhere else in the article or page being consulted, or from another source) results in the expansion of a query, thereby **supporting learning and understanding**. Adding an entity to the query from the result set **enables facets and metadata-based result filtering**.

Entities become resources for users to express their information interests and search intents. Sets of related or previously observed entities can be used to collect feedback from users on their current reliance, which would support advanced personalization in iterative user modeling where the exploration system presents predictions of user intent through sets of entities helping the user to discover and formulate her current intent.

Modern browsers and operating systems already implement affordances to inspect the definition of an expression, its corresponding Wikipedia entry or related search engine results. Such affordance is here generalized, using entity search to yield a crop of related entities from any selected object (e.g., an expression, article, or link).

Affordance 2: Entity Mapping

Entities can be moved around, and users are provided with the spatial freedom to organize the entities of interest in a layout that reflects their understanding and their mental representation of the information space.

Spatial organization of thoughts is a common behavior. We draw mind maps, we make piles of documents, we organize sticky notes, and store documents within directories or under consistent tags. Sense-making is an important part of exploratory search [White and Roth, 2009], and as such it relies on users building a mental representation of the state of the world (i.e., the information space at hand) and then iteratively contrasting this representation against the real world (i.e., new information) to update it and acquire a progressively more accurate understanding of the information space [Russell et al., 1993].

Entity mapping provides support for mind mapping, which **supports learning and understanding**. It provides an implicit input channel for **leveraging the search context**. It also allows for creating **visualizations that support insight and decision-making** by enabling multi-aspect search, as well as for addressing the need for **histories, workspaces, and progress updates**.

Affordance 3: Entity Storing and Sharing

Entities and groups of entities can be easily saved for later use and shared with collaborators.

In today's paradigm, documents often serve as units of information. Users search for, bookmark, and share such documents. Such actions are not sufficient, however. The user often forgets the intent behind the bookmarking and thus loses the utility of the stored document. Some additional action is required, such as giving each bookmark a context-relevant title or organizing bookmarks within theme-specific directories. Sharing requires the use of messaging channels, as text messages are necessary to convey context and intent. Entity-oriented information enables the use of variable and personalized units of information. Users can search for, store, and share references to persons, media, excerpts, and organizations. Taking advantage of affordances 1 and 2, the exchange of information involves potentially sharing – and collaborating on – whole contexts in the form of organized entities, which **facilitates collaboration**. The same principle gives access to these contexts across devices, providing flexible **support for task management** and **enabling histories, workspaces, and progress updates**. Stored or saved information also provides an implicit input channel for **leveraging the search context**.

The present template consists of fundamental principles aimed at guiding the design of future systems and supporting information exploration while also limiting the number of constraints imposed on the overall design space. In this section, we discuss aspects that are not addressed by the template and attempt to outline the remaining design space.

Hypercues are designed to be identified and defined by users (although they could also be recommended within contents). For instance, in the latest iteration of its operating system for tablets (iOS 11), Apple has introduced a generalized ability to drag and drop. Pictures, text snippets, news articles, hyperlinks, and other bit of information pop out of the environment with a gesture of the finger, thus becoming interactive objects that can be dragged across applications and dropped into messages, notes, or cloud-based storage. This ability lets the user interact with predefined objects and with user-defined selections, and it offers an ideal interactive base for the integration of the affordances proposed in this paper.

Although the template does not provide information about the shape and size of displayed hypercues, it is useful to discuss the requirements and provide some recommendations based on our experience. The first requirements of the hypercue marker is for the represented entity to be identifiable and placed in a space that allows it to be moved and positioned in relation to other entities. A constant challenge when designing entity-based interfaces balancing the amount of information conveyed through the entity marker against the number of entities that can be comfortably displayed. In any case, it is necessary to provide the option to quickly inspect the entity, so that the user can access a comprehensive overview of the entity through linked content and related material. However, it is also essential to show enough information up-front to trigger the user's recognition and incite her interest. Modern desktop-based operating systems offer a good model for representing files and directories as manipulable objects using an icon and one or two short lines of text. The most useful information depends on the task and on the information space. Although movies are usually displayed with a poster, a title and a release year, finding the most relevant movie in a set could depend on other information, such as the cast or the rating. Likewise, finding useful academic articles can require variable criteria (e.g., authors, venue or citations). The solution might lie in a balance between user-defined preferences and automated, context-sensitive, and adaptive interface settings.

These guidelines are generalizable to every information space. Their reliance on direct manipulation and spatial layouts makes the hypercue a potentially interesting candidate for integration with the physical world through playful tangible interactions. Registering an entity or a set of entities as physical objects allows users to combine and share such objects to playfully discover information through machine vision or sensing surfaces.

7.3 Implications of the Research

An important implication derives from the fact that all interactions explored in this work require substantially more effort from the user than current methods require, as users have grown accustomed to content feeds and to the simplicity and immediacy of today's search engines. I advocate for information practices in which users are more active, and posit that this is the cost of providing greater transparency and control over information. However, this cost can be mitigated through fluidity by having every interaction serve an informational goal and letting the user become truly absorbed by the task, thus rewarding her with persistent and constructive search sessions that remain useful in the long run.

We can imagine that, in the future, a user's search interface of choice will be independent from the data being explored. As a result, a user could apply her tool of choice to discover information of interest indifferently within academic articles, music, movies, and social media posts, thus increasing the potential for serendipity and creative solutions. A standardized base set of affordances such as the Hypercue provides not only a useful design template for the future design of such tools, as well as establishing a consistent and effective paradigm in the way we will, as users, interact with information.

Beyond the purely interactive layer of potential information seeking techniques discussed in this thesis, it is important to remember that the present research presumes the existence of organized entities, and the road to such reality most likely involves machine learning and artificial intelligence constructing such entities from the current amorphous text-based information that populates our searchable information space. As I conveniently delegate, in my introductory chapter, all responsibility related to the organization of information to the *semantic web* and *information retrieval* communities, I must acknowledge that such task is not only difficult but harbors the very same potential lack of transparency that the present research addresses in current search systems.

7.4 Limitations and Future Directions

The lack of readily available entity-oriented information has made this research a challenging venture. As a consequence, our various prototypes and experiments relied on limited closed – although very large – sets of home-indexed entity-oriented data. For that reason, our studies were limited to the laboratory, under controlled conditions. Moreover, our focus on fundamental aspects of entity interaction encouraged systematic comparison with baseline conditions replicating conventional methods. This approach yielded a strong and focused understanding of the benefits of our designs over the *de facto* paradigm, and I believe was the right strategy at this point in the research on this topic. But an important trade-off results from the fact that for most experimental tasks, participants were driven by extrinsic motivations, i.e., topics of interests that were provided to them, instead of intrinsic motivation in the form of free exploration following their own inspiration. Another limitation of our comparative studies lies in the minute differences between our baseline systems and the actual system they imitate. While great care has been taken in creating these systems, their models, e.g. Google Search, incorporate beneficial features, e.g. predictive typing based on a large number of queries, that could not be replicated at the time. In all cases, the experimental system was treated comparably, as it also was missing such features. A natural future research direction would consist of utilizing affordances proposed in this research to implement systems that are ready for larger studies in the field. Such a setting, coupled with qualitative research methods, would yield a richer understanding of potential futures of how we, as users, will interact with information, and how future paradigms could change and benefit our societies.

References

[Abras et al., 2004] Abras, C., Maloney-Krichmar, D., Preece, J., et al. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4):445–456.

[Ahlberg and Shneiderman, 1994] Ahlberg, C. and Shneiderman, B. (1994). Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. CHI'94*, pages 313–317. ACM.

[Ahn and Brusilovsky, 2009] Ahn, J.-W. and Brusilovsky, P. (2009). Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization*, 8(3):167–179.

[Alsallakh et al., 2014] Alsallakh, B., Micallef, L., Aigner, W., Hauser, H., Miksch, S., and Rodgers, P. (2014). Visualizing sets and set-typed data: State-of-the-art and future challenges. In *EuroVis– State of The Art Reports*, pages 1–21. Eurographics.

[Amento et al., 1999] Amento, B., Hill, W., Terveen, L., Hix, D., and Ju, P. (1999). An empirical evaluation of user interfaces for topic management of web sites. In *Proc. CHI'99*, pages 552–559. ACM.

[Andolina et al., 2019] Andolina, S., Daee, P., Vuong, T., Ruotsalo, T., Klouche, K., Sjöberg, M., Kaski, S., and Jacucci, G. (2019). Proactive recommendation in context: From relevant items to actionable entities. *ACM Transactions on Computer-Human Interaction*.

[Andolina et al., 2015a] Andolina, S., Klouche, K., Cabral, D., Ruotsalo, T., and Jacucci, G. (2015a). Inspirationwall: supporting idea generation through automatic information exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pages 103–106. ACM.

[Andolina et al., 2015b] Andolina, S., Klouche, K., Peltonen, J., Hoque, M., Ruotsalo, T., Cabral, D., Klami, A., Głowacka, D., Floréen, P., and Jacucci, G. (2015b). Intentstreams: smart parallel search streams for branching exploratory search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 300–305. ACM.

[Andolina et al., 2018a] Andolina, S., Klouche, K., Ruotsalo, T., Floréen, P., and Jacucci, G. (2018a). Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing & Management*, 54(6):1182–1202.

[Andolina et al., 2018b] Andolina, S., Orso, V., Schneider, H., Klouche, K., Ruotsalo, T., Gamberini, L., and Jacucci, G. (2018b). Investigating proactive search support in conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference*, pages 1295–1307. ACM.

[Balasubramanian and Cucerzan, 2010] Balasubramanian, N. and Cucerzan, S. (2010). Topic pages: An alternative to the ten blue links. In *2010 IEEE Fourth International Conference on Semantic Computing*, pages 353–360.

[Balog, 2018] Balog, K. (2018). *Entity-oriented search*. Springer.

[Balog et al., 2010] Balog, K., Meij, E., and De Rijke, M. (2010). Entity search: building bridges between two worlds. In *Proceedings of the 3rd International Semantic Search Workshop*, page 9. ACM.

[Barral et al., 2018] Barral, O. et al. (2018). *Implicit Interaction with Textual Information using Physiological Signals*. PhD thesis, Helsingin yliopisto.

[Bates, 1989] Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5):407–424.

[Benford et al., 1995] Benford, S., Snowdon, D., Greenhalgh, C., Ingram, R., Knox, I., and Brown, C. (1995). Vr-vibe: A virtual environment for co-operative information retrieval. In *Computer Graphics Forum*, volume 14, pages 349–360. Wiley.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, 284(5):28–37.

[Bier et al., 2008] Bier, E. A., Card, S. K., and Bodnar, J. W. (2008). Entity-based collaboration tools for intelligence analysis. In *2008 IEEE Symposium on Visual Analytics Science and Technology*, pages 99–106. IEEE.

[Boyack et al., 2002] Boyack, K. W., Wylie, B. N., and Davidson, G. S. (2002). Domain visualization using VxInsight® for science and technology management. *JASIST*, 53(9):764–774.

[Brooke et al., 1996] Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

[Brown et al., 2015] Brown, B., McGregor, M., and McMillan, D. (2015). Searchable objects: Search in everyday conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 508–517, New York, NY, USA. ACM.

[Capra et al., 2007] Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., and Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 442–451. ACM.

[Carmel et al., 2012] Carmel, D., Zwerdling, N., and Yogev, S. (2012). Entity oriented search and exploration for cultural heritage collections: the eu cultura project. In *Proceedings of the 21st International Conference on World Wide Web*, pages 227–230. Citeseer.

[Carrol, 1999] Carrol, J. M. (1999). Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, pages 11–pp. IEEE.

[Chalmers and Chitson, 1992] Chalmers, M. and Chitson, P. (1992). Bead: Explorations in information visualization. In *Proc. SIGIR'92*, pages 330–337. ACM.

[Chau et al., 2011] Chau, D. H., Kittur, A., Hong, J. I., and Faloutsos, C. (2011). Apolo: making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 167–176. ACM.

[Chaudhry et al., 2015] Chaudhry, A., Crowcroft, J., Howard, H., Madhavapeddy, A., Mortier, R., Haddadi, H., and McAuley, D. (2015). Personal data: Thinking inside the box. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives*, AA '15, pages 29–32. Aarhus University Press.

[Clarkson et al., 2009] Clarkson, E., Desai, K., and Foley, J. (2009). Resultmaps: Visualization for search interfaces. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1057–1064.

[Das-Neves et al., 2005] Das-Neves, F., Fox, E. A., and Yu, X. (2005). Connecting topics in document collections with stepping stones and pathways. In *Proc. CIKM'05*, pages 91–98. ACM.

[de Montjoye et al., 2014] de Montjoye, Y.-A., Shmueli, E., Wang, S. S., and Pentland, A. S. (2014). openPDS: protecting the privacy of metadata through SafeAnswers. *PloS one*, 9(7).

[Deerwester et al., 1990] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

[Dewan and Choudhary, 1991] Dewan, P. and Choudhary, R. (1991). Flexible user interface coupling in a collaborative system. In *1991 SIGCHI Conference on Human Factors in Computing Systems, CHI 1991*. Association for Computing Machinery.

[Dörk et al., 2008] Dörk, M., Carpendale, S., Collins, C., and Williamson, C. (2008). Visgets: Coordinated visualizations for web-based information exploration and discovery. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1205–1212.

[Dörk et al., 2011] Dörk, M., Carpendale, S., and Williamson, C. (2011). The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1215–1224. ACM.

[Elmqvist et al., 2011] Elmqvist, N., Moere, A. V., Jetter, H.-C., Cernea, D., Reiterer, H., and Jankun-Kelly, T. (2011). Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340.

[Esenther, 2006] Esenther, F. D. (2006). Better mouse support for touch-based interactions.

[Findeli et al., 2008] Findeli, A., Brouillet, D., Martin, S., Moineau, C., and Tarrago, R. (2008). Research through design and transdisciplinarity: A tentative contribution to the methodology of design research. In *Swiss Design network symposium*, page 67.

[Fortuna et al., 2005] Fortuna, B., Grobelnik, M., and Mladenic, D. (2005). Visualization of text document corpus. *Informatica*, 29(4):497–502.

[Garigliotti and Balog, 2018] Garigliotti, D. and Balog, K. (2018). Towards an understanding of entity-oriented search intents. In *European Conference on Information Retrieval*, pages 644–650. Springer.

[Glowacka et al., 2013] Glowacka, D., Ruotsalo, T., Konuyshkova, K., Kaski, S., Jacucci, G., et al. (2013). Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, pages 117–128. ACM.

[González-Ibáñez et al., 2013] González-Ibáñez, R., Haseki, M., and Shah, C. (2013). 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.

[Gratzl et al., 2013] Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., and Streit, M. (2013). Lineup: Visual analysis of multi-attribute rankings. *TVCG*, 19(12):2277–2286.

[Guex, 2016] Guex, G. (2016). *Réseaux spatiaux: modèles, classification et flux*. PhD thesis, Université de Lausanne, Faculté des géosciences et de l'environnement.

[Guimbretière et al., 2001] Guimbretière, F., Stone, M., and Winograd, T. (2001). Fluid interaction with high-resolution wall-size displays. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 21–30. ACM.

[Gutwin and Greenberg, 2002] Gutwin, C. and Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11(3-4):411–446.

[Gyllstrom and Soules, 2008] Gyllstrom, K. and Soules, C. (2008). Seeing is retrieving: Building information context from what the user sees. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, IUI '08, pages 189–198, New York, NY, USA. ACM.

[Havre et al., 2001] Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. (2001). Interactive visualization of multiple query results. In *Information Visualization*, page 105. IEEE.

[Hearst, 1995] Hearst, M. A. (1995). Tilebars: Visualization of term distribution information in full text information access. In *Proc. CHI'95*, pages 59–66. ACM.

[Hearst, 2006] Hearst, M. A. (2006). Clustering versus faceted categories for information exploration. *Communications of the ACM*, 49(4):59–61.

[Hearst, 2014] Hearst, M. A. (2014). What's missing from collaborative search? *Computer*, 47(3):58–61.

[Hetzler and Turner, 2004] Hetzler, E. and Turner, A. (2004). Analysis experiences using information visualization. *IEEE CG&A*, 24(5):22–26.

[Hinrichs and Forlini, 2017] Hinrichs, U. and Forlini, S. (2017). In defense of sandcastles: Research thinking through visualization in dh. In *Proceedings of the conference on Digital Humanities*. International Alliance of Digital Humanities Organizations (ADHO).

[Hoeber and Yang, 2006a] Hoeber, O. and Yang, X. D. (2006a). Interactive web information retrieval using wordbars. In *International Conference on Web Intelligence*, pages 875–882. IEEE.

[Hoeber and Yang, 2006b] Hoeber, O. and Yang, X. D. (2006b). The visual exploration of web search results using hotmap. In *Proc. IV'06*, pages 157–165. IEEE Computer Society.

[Hutchins et al., 1985] Hutchins, E. L., Hollan, J. D., and Norman, D. A. (1985). Direct manipulation interfaces. *Human-computer interaction*, 1(4):311–338.

[Jansen and Spink, 2006] Jansen, B. J. and Spink, A. (2006). How are we searching the world wide web? a comparison of nine search engine transaction logs. *Information processing & management*, 42(1):248–263.

[Johnson and Shneiderman, 1991] Johnson, B. and Shneiderman, B. (1991). *Treemaps: A space-filling approach to the visualization of hierarchical information structures*. IEEE.

[Kang et al., 2012] Kang, C., Wang, X., Chang, Y., and Tseng, B. (2012). Learning to rank with multi-aspect relevance for vertical search. In *Proc. WSDM'12*, pages 453–462. ACM.

[Kaski et al., 1998] Kaski, S., Honkela, T., Lagus, K., and Kohonen, T. (1998). WEBSOM–self-organizing maps of document collections. *Neurocomputing*, 21(1):101–117.

[Kerne et al., 2006] Kerne, A., Koh, E., Dworaczyk, B., Mistrot, J. M., Choi, H., Smith, S. M., Graeber, R., Caruso, D., Webb, A., Hill, R., and Albea, J. (2006). combinformation: A mixed-initiative system for representing collections as compositions of image and text surrogates. In *Proc. JCDL'06*, pages 11–20. ACM.

[Kerne and Smith, 2004] Kerne, A. and Smith, S. M. (2004). The information discovery framework. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 357–360. ACM.

[Kidwell et al., 2008] Kidwell, P., Lebanon, G., and Cleveland, W. S. (2008). Visualizing incomplete and partially ranked data. *TVCG*, 14(6):1356–1363.

[Kilgour et al., 2010] Kilgour, J., Carletta, J., and Renals, S. (2010). The ambient spotlight: Queryless desktop search from meeting speech. In *Proceedings of the 2010 International Workshop on Searching Spontaneous Conversational Speech*, SSCS '10, pages 49–52, New York, NY, USA. ACM.

[Kirsh, 2014] Kirsh, D. (2014). The importance of chance and interactivity in creativity. *Pragmatics & Cognition*, 22(1):5–26.

[Klein et al., 2006] Klein, G., Moon, B., and Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE intelligent systems*, 21(4):70–73.

[Klouche et al., 2015] Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., and Jacucci, G. (2015). Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4189–4198. ACM.

[Klouche et al., 2018] Klouche, K., Ruotsalo, T., and Jacucci, G. (2018). From hyperlinks to hypercues: Entity-based affordances for fluid information exploration. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*, pages 401–411. ACM.

[Klouche et al., 2017] Klouche, K., Ruotsalo, T., Micallef, L., Andolina, S., and Jacucci, G. (2017). Visual re-ranking for multi-aspect information retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 57–66. ACM.

[König and Rasch, 2014] König, R. and Rasch, M. (2014). *Society of the Query Reader: Reflections on Web Search.*

[Kucher and Kerren, 2015] Kucher, K. and Kerren, A. (2015). Text visualization techniques: Taxonomy, visual survey, and community insights. In *PacificVis'15*, pages 117–121. IEEE Computer Society.

[Kules and Shneiderman, 2008] Kules, B. and Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. *IPM*, 44(2):463–484.

[Lee et al., 2009] Lee, B., Smith, G., Robertson, G. G., Czerwinski, M., and Tan, D. S. (2009). Facetlens: exposing trends and relationships to support sensemaking within faceted datasets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1293–1302. ACM.

[Lewandowski, 2014] Lewandowski, D. (2014). Why we need an independent index of the web. *arXiv preprint arXiv:1405.2212*.

[Lewandowski, 2015] Lewandowski, D. (2015). Living in a world of biased search engines. *Online Information Review*, 39(3).

[Lin, 1997] Lin, X. (1997). Map displays for information retrieval. *JASIS*, 48(1):40–54.

[Mackinlay, 1986] Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *TOG*, 5(2):110–141.

[Marchionini, 2006] Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46.

[Matejka et al., 2015] Matejka, J., Anderson, F., and Fitzmaurice, G. (2015). Dynamic opacity optimization for scatter plots. In *Proc. CHI'15*, pages 2707–2710. ACM.

[McGrath, 1995] McGrath, J. E. (1995). Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human–Computer Interaction*, pages 152–169. Elsevier.

[McGregor and Tang, 2017] McGregor, M. and Tang, J. C. (2017). More to meetings: Challenges in using speech-based technology to support meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 2208–2220, New York, NY, USA. ACM.

[McMillan et al., 2015] McMillan, D., Loriette, A., and Brown, B. (2015). Repurposing conversation: Experiments with the continuous speech stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 3953–3962, New York, NY, USA. ACM.

[McTear, 2002] McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)*, 34(1):90–169.

[Miliaraki et al., 2015] Miliaraki, I., Blanco, R., and Lalmas, M. (2015). From selena gomez to marlon brando: Understanding explorative entity search. In *Proceedings of the 24th International Conference on World Wide Web*, pages 765–775. International World Wide Web Conferences Steering Committee.

[Miller et al., 1998] Miller, N. E., Wong, P. C., Brewster, M., and Foote, H. (1998). Topic islands tm-a wavelet-based text visualization system. In *Proc. VIS*, pages 189–196. IEEE.

[Morris et al., 2010] Morris, M. R., Fisher, D., and Wigdor, D. (2010). Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. *Information processing & management*, 46(6):703–717.

[Morse and Lewis, 1997] Morse, E. and Lewis, M. (1997). Why information retrieval visualizations sometimes fail. In *Proc. IEEE SMC'97*, volume 2, pages 1680–1685.

[Munzner et al., 2003] Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L., and Zhou, Y. (2003). TreeJuxtaposer: scalable tree comparison using focus + context with guaranteed visibility. *TOG*, 22(3):453–462.

[Nakamura and Csikszentmihalyi, 2014] Nakamura, J. and Csikszentmihalyi, M. (2014). The concept of flow. In *Flow and the foundations of positive psychology*, pages 239–263. Springer.

[Negri et al., 2014] Negri, M., Turchi, M., de Souza, J. G. C., and Falavigna, D. (2014). Quality estimation for automatic speech recognition. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*, pages 1813–1823.

[Norman, 2013] Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

[Nuchprayoon and Korfhage, 1994] Nuchprayoon, A. and Korfhage, R. R. (1994). Guido, a visual tool for retrieving documents. In *Proc. VL/HCC'94*, pages 64–71. IEEE.

[O'Brien and Toms, 2010] O'Brien, H. L. and Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69.

[Olsen et al., 1993] Olsen, K. A., Korfhage, R. R., Sochats, K. M., Spring, M. B., and Williams, J. G. (1993). Visualization of a document collection: The VIBE system. *IPM*, 29(1):69–81.

[O'brien et al., 2014] O'brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., and Cook, D. A. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine*, 89(9):1245–1251.

[O'Brien et al., 2018] O'Brien, H. L., Cairns, P., and Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112:28–39.

[O'brien and Toms, 2013] O'brien, H. L. and Toms, E. G. (2013). Examining the generalizability of the user engagement scale (ues) in exploratory search. *Information Processing & Management*, 49(5):1092–1107.

[Pariser, 2011] Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

[Pirolli and Card, 1999] Pirolli, P. and Card, S. (1999). Information foraging. *Psychological review*, 106(4):643.

[Pirolli and Card, 2005] Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, volume 5, pages 2–4. McLean, VA, USA.

[Pound et al., 2010] Pound, J., Mika, P., and Zaragoza, H. (2010). Ad-hoc object retrieval in the web of data. In *Proceedings of the 19th international conference on World wide web*, pages 771–780. ACM.

[Pu et al., 2011] Pu, P., Chen, L., and Hu, R. (2011). A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 157–164. ACM.

[Raghavan and Wong, 1986] Raghavan, V. V. and Wong, S. M. (1986). A critical analysis of vector space model for information retrieval. *JASIS*, 37(5):279–287.

[Ramos and Balakrishnan, 2003] Ramos, G. and Balakrishnan, R. (2003). Fluid interaction techniques for the control and annotation of digital video. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 105–114. ACM.

[Rochat and Triclot, 2017] Rochat, Y. and Triclot, M. (2017). The networks of science fiction characters: intermediate reading samples. *Res Futurae - Revue d'études sur la Science-Fiction*, (10).

[Rohrer et al., 1998] Rohrer, R. M., Ebert, D. S., and Sibert, J. L. (1998). The shape of Shakespeare: visualizing text using implicit surfaces. In *Proc. InfoVis*, pages 121–129. IEEE.

[Ruotsalo et al., 2013a] Ruotsalo, T., Athukorala, K., Głowacka, D., Konyushkova, K., Oulasvirta, A., Kaipiainen, S., Kaski, S., and Jacucci, G. (2013a). Supporting exploratory search tasks with interactive user modeling. In *Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries*, page 39. American Society for Information Science.

[Ruotsalo et al., 2015] Ruotsalo, T., Jacucci, G., Myllymäki, P., and Kaski, S. (2015). Interactive intent modeling: Information discovery beyond search. *Communications of the ACM*, 58(1):86–92.

[Ruotsalo et al., 2016] Ruotsalo, T., Klouche, K., Cabral, D., Andolina, S., and Jacucci, G. (2016). Flexible entity search on surfaces. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, pages 175–179. ACM.

[Ruotsalo et al., 2013b] Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K., Kosunen, I., Reijonen, A., Myllymäki, P., Jacucci, G., et al. (2013b). Directing exploratory search with interactive intent modeling. In *Proc. CIKM'13*, pages 1759–1764. ACM.

[Russell et al., 1993] Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K. (1993). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 269–276. ACM.

[Sanders and Westerlund, 2011] Sanders, E. B.-N. and Westerlund, B. (2011). Experiencing, exploring and experimenting in and with co-design spaces. *Nordes*, (4).

[Scott et al., 2004] Scott, S. D., Carpendale, M. S. T., and Inkpen, K. M. (2004). Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 294–303. ACM.

[Seo and Shneiderman, 2005] Seo, J. and Shneiderman, B. (2005). A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2):96–113.

[Shi et al., 2012] Shi, C., Cui, W., Liu, S., Xu, P., Chen, W., and Qu, H. (2012). RankExplorer: Visualization of ranking changes in large time series data. *TVCG*, 18(12):2669–2678.

[Shi et al., 2017] Shi, Y., Wang, Y., Qi, Y., Chen, J., Xu, X., and Ma, K.-L. (2017). Ideawall: Improving creative collaboration through combinatorial visual stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17, pages 594–603, New York, NY, USA. ACM.

[Shneiderman, 1993] Shneiderman, B. (1993). 1.1 direct manipulation: a step beyond programming languages. *Sparks of innovation in human-computer interaction*, 17:1993.

[Shneiderman, 1994] Shneiderman, B. (1994). Dynamic queries for visual information seeking. *Software*, 11(6):70–77.

[Shneiderman, 1995] Shneiderman, B. (1995). Looking for the bright side of user interface agents. *interactions*, 2(1):13–15.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. VL/HCC'96*, pages 336–343. IEEE.

[Shneiderman, 1997] Shneiderman, B. (1997). Direct manipulation for comprehensible, predictable and controllable user interfaces. In *IUI*, volume 97, pages 33–39.

[Shneiderman et al., 2000] Shneiderman, B., Feldman, D., Rose, A., and Grau, X. F. (2000). Visualizing digital library search results with categorical and hierarchical axes. In *Proc. DL'00*, pages 57–66. ACM.

[Shneiderman et al., 2009] Shneiderman, B., Plaisant, C., Cohen, M., and Jacobs, S. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, 5th edition.

[Singh et al., 2011] Singh, R. I., Sumeeth, M., and Miller, J. (2011). Evaluating the readability of privacy policies in mobile environments. *International Journal of Mobile Human Computer Interaction (IJMHCI)*, 3(1):55–78.

[Sjöberg et al., 2017] Sjöberg, M., Chen, H.-H., Floréen, P., Koskela, M., Kuikkaniemi, K., Lehtiniemi, T., and Peltonen, J. (2017). Digital me: Controlling and making sense of my digital footprint. In *Symbiotic Interaction*, Lecture Notes in Computer Science, pages 155–167. Springer.

[Smith et al., 2006] Smith, G., Czerwinski, M., Meyers, B., Robbins, D., Robertson, G., and Tan, D. S. (2006). Facetmap: A scalable search and browse visualization. *IEEE Transactions on visualization and computer graphics*, 12(5):797–804.

[Song et al., 2012] Song, W., Yu, Q., Xu, Z., Liu, T., Li, S., and Wen, J.-R. (2012). Multi-aspect query summarization by composite query. In *Proc. SIGIR'12*, pages 325–334. ACM.

[Spoerri, 1993] Spoerri, A. (1993). InfoCrystal: A visual tool for information retrieval & management. In *Proc. CIKM'93*, pages 11–20. ACM.

[Stasko et al., 2008] Stasko, J., Görg, C., and Liu, Z. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132.

[Stefik et al., 1987] Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., and Suchman, L. (1987). Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Communications of the ACM*, 30(1):32–48.

[Stolte et al., 2002] Stolte, C., Tang, D., and Hanrahan, P. (2002). Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):52–65.

[Suchman, 2007] Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

[Tang and Joiner, 2006] Tang, Y. and Joiner, C. (2006). *Synergic inquiry: A collaborative action methodology*. Sage.

[Teevan et al., 2005] Teevan, J., Dumais, S. T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456. ACM.

[Thudt et al., 2012] Thudt, A., Hinrichs, U., and Carpendale, S. (2012). The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1461–1470. ACM.

[Thudt et al., 2015] Thudt, A., Hinrichs, U., and Carpendale, S. (2015). A modular approach to promote creativity and inspiration in search. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, pages 245–254. ACM.

[Varcholik et al., 2012] Varcholik, P. D., Laviola Jr, J. J., and Hughes, C. E. (2012). Establishing a baseline for text entry for a multi-touch virtual keyboard. *International Journal of Human-Computer Studies*, 70(10):657–672.

[Vuong et al., 2017a] Vuong, T., Jacucci, G., and Ruotsalo, T. (2017a). Proactive information retrieval via screen surveillance. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, pages 1313–1316, New York, NY, USA. ACM.

[Vuong et al., 2017b] Vuong, T., Jacucci, G., and Ruotsalo, T. (2017b). Watching inside the screen: Digital activity monitoring for task recognition and proactive information retrieval. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):109:1–109:23.

[Ware, 2012] Ware, C. (2012). *Information visualization: perception for design*. Elsevier.

[White and Morris, 2007] White, R. W. and Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 255–262. ACM.

[White and Roth, 2009] White, R. W. and Roth, R. A. (2009). Exploratory search: beyond the query-response paradigm (synthesis lectures on information concepts, retrieval & services). *Morgan and Claypool Publishers*, 3.

[Wise et al., 1995] Wise, J., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., Crow, V., et al. (1995). Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proc. Information Visualization*, pages 51–58. IEEE.

[Xin et al., 2006] Xin, D., Han, J., Cheng, H., and Li, X. (2006). Answering top-k queries with multi-dimensional selections: The ranking cube approach. In *Proc. VLDB'06*, pages 463–474. VLDB.

[Yang et al., 2015] Yang, D., Zhang, D., Chen, L., and Qu, B. (2015). Nationtelescope: Monitoring and visualizing large-scale collective behavior in lbsns. *Journal of Network and Computer Applications*, 55:170 – 180.

[Zhai et al., 2015] Zhai, C., Cohen, W. W., and Lafferty, J. (2015). Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *ACM SIGIR Forum*, volume 49, pages 2–9. ACM.

[Zhang, 2001] Zhang, J. (2001). Tofir: A tool of facilitating information retrieval–introduce a visual retrieval model. *IPM*, 37(4):639–657.

[Zhang and Korfhage, 1999] Zhang, J. and Korfhage, R. R. (1999). Dare: Distance and angle retrieval environment: A tale of the two measures. *JASIS*, 50(9):779–787.

[Zhu et al., 2013] Zhu, Z., Blanke, U., Calatroni, A., and Tröster, G. (2013). Human activity recognition using social media data. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, MUM '13, pages 21:1–21:10, New York, NY, USA. ACM.

[Zimmerman et al., 2007] Zimmerman, J., Forlizzi, J., and Evenson, S. (2007). Research through design as a method for interaction design research in hci. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 493–502. ACM.

Publication I

Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. Designing for Exploratory Search on Touch Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, (CHI '15). New York, NY, USA, pp. 4189-4198, ACM. DOI: 10.1145/2702123.2702489, April 2015.

Designing for Exploratory Search on Touch Devices

Khalil Klouche [1,3], **Tuukka Ruotsalo** [2], **Diogo Cabral** [1]
Salvatore Andolina [1], **Andrea Bellucci** [4], **Giulio Jacucci** [1,2]
[1] Helsinki Institute for Information Technology HIIT, Department of Computer Science,
University of Helsinki, PO Box 68, 00014 University of Helsinki, Finland
[2] Helsinki Institute for Information Technology HIIT, Aalto University,
PO Box 15600, 00076 Aalto, Finland
[3] Aalto University, School of Arts, Design and Architecture, Media Lab Helsinki,
Hameentie 135 c, 00560 Helsinki, Finland
[4] Universidad Carlos III de Madrid, Av. Universidad 30, 28911 Leganes (Madrid), Spain
[1] first.last@helsinki.fi, [2] first.last@aalto.fi, [4] abellucc@inf.uc3m.es

ABSTRACT
Exploratory search confront users with challenges in expressing search intents as the current search interfaces require investigating result listings to identify search directions, iterative typing, and reformulating queries. We present the design of Exploration Wall, a touch-based search user interface that allows incremental exploration and sense-making of large information spaces by combining entity search, flexible use of result entities as query parameters, and spatial configuration of search streams that are visualized for interaction. Entities can be flexibly reused to modify and create new search streams, and manipulated to inspect their relationships with other entities. Data comprising of task-based experiments comparing Exploration Wall with conventional search user interface indicate that Exploration Wall achieves significantly improved recall for exploratory search tasks while preserving precision. Subjective feedback supports our design choices and indicates improved user satisfaction and engagement. Our findings can help to design user interfaces that can effectively support exploratory search on touch devices.

Author Keywords
User Interfaces; Exploratory Search; Touch Devices

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces; H.3.3. Information Storage and Retrieval: Information Search and Retrieval

CHI 2014, April 18 - 23, 2015, Seoul, Republic of Korea

http://dx.doi.org/10.1145/2702123.2702489

INTRODUCTION
Surface computing technologies hold great potential for enhancing information retrieval activities. Devices with touch interaction capabilities make possible to design engaging direct manipulation interactions, facilitate awareness of information available for the user beyond conventional search engine result pages, and afford visualization and spatial organization of content. However, conventional search user interfaces rely exclusively on typed-query interaction and result presentation as ranked list of documents [24], and thus they present challenges when transferred to touch devices.

These design conventions are pernicious for many search scenarios, in particular, *exploratory search* scenarios, that describe a class of search activities that go beyond basic lookup, typically involving the user in a field with which she is not familiar [38]. Exploratory search patterns are very diverse, but share together common complex user-centered challenges: The need to overcome difficulties in formulating queries in unknown information spaces, ways to learn about the information space and identify possible search directions beyond the entry point specified by an initial query [28].

Causes of inadequacy of classical search interfaces for touch-enabled devices are the poor substitutes for keyboard and mouse inputs. Virtual keyboards are reported to be less performing then their physical counterpart [34], and do not provide usual text editing shortcuts (e.g., copy, cut, paste, cancel). As for mouse-based interactions, touch-based substitutes constrain natural touch interactions and prove difficult for quick and accurate text selection [10]. Also, the lack of window management on touch devices does only allow the visualization of a single query at a time, which hinders comparison and revisiting previously retrieved information.

As a consequence, the fluidity and search performance expected while searching with surfaces is hampered [38]. Therefore, it becomes crucial to design new solutions that overcome the limitations of conventional search user interfaces and bring in forefront the potential of multi-touch interaction.

We present the design of Exploration Wall, a novel search user interface for facilitating exploratory search tasks on touch devices. Exploration Wall is based on the following design principles targeting above-mentioned challenges:

1. Flexible reuse and combination of items to facilitate query formulation.

2. Result sets of not only documents but most relevant entities to foster iterative query reformulation.

3. Use of spatial configuration of multiple search streams to identify search directions and learn about the information space.

Our design was found to facilitate exploratory search behavior when compared to the conventional baseline search user interface, as indicated by measured system effectiveness. Moreover, users were found to be more engaged with the task and subjectively more satisfied by their exploratory search. Our findings suggest that our principles can be effective when designing search user interfaces for touch devices, and can overcome many limitations of the direct adaptation of conventional search user interfaces to surfaces.

BACKGROUND

Exploratory Search

Most available tools for information retrieval focus on lookup retrieval, such as looking up the address of a restaurant or reminding a historical fact, while many users search to solve more complex tasks that require exploration of the information space. White [38] describes exploratory search as activities that move beyond basic lookup retrieval. Such activities rely on learning and investigation [22]. Exploratory search activities have no predetermined goals and are described as open-ended [38]. Therefore, the absence of clear user intents leads to difficulties in formulating queries.

Exploratory search processes are considered dynamic. As the information space is unknown or unfamiliar to the user, the query formulation evolves iteratively as the user becomes more familiar with the context [7]. In addition, exploratory search tasks include cognitive and behavioral attributes [40]. Cognitive attributes can be defined as those that have the reasoning associated with conducting an exploratory search and involve learning and investigation as goals; general and ill-structured problems; uncertainty; dynamic, multi-faceted and complex search tasks and are accompanied with sensemaking, decision making or other cognition.

As stated, in the context of exploratory search users need particular support in formulating queries, learning about the information space and identifying possible search directions. Next we review how visual interfaces have been developed to address these issues.

Visual Interfaces in Search

Recently, a variety of search systems have been developed in order to enable faster relevance judgment and effective feedback [12, 17, 19, 23, 32]. Several visualization approaches have been explored including multiple linked lists, scatter plots, graphs and their combinations [31, 19]. These types of visual search systems are distinguished from familiar query composition ones (e.g., Project Blacklight [30]) because of their emphasis on rapid filtering to reduce result sets, progressive refinement of search parameters, continuous reformulation of goals, and visual scanning to identify results [1].

Currently, visual approaches attempt to better support exploration in different ways: supporting sense making by incrementally and interactively exploring the network of data [8], showing how visualization support user involvement in the recommendation providing rationale behind suggested items [35], visualizing relations of different queries and result sets [2]. Recent work shows how to support users to view and manipulate their search intent models as viewed by the system [2, 3, 29, 11] . This work attempts to combine personalization of search with visualization approaches offering support to formulate queries and learning about the information space helping users in directing their search [29] . While these systems demonstrate the importance of investigating visual user interfaces in exploratory search, they have not been considered for multi-touch devices.

Recent visual interfaces in search have shown the effectiveness of interacting visually with query elements and results combined with computational techniques that support exploration. Multi-touch devices could provide an opportunity for visual interactive search for their capability to encourage manipulation of visual elements and for the limits posed by the absence of mouse and keyboard. In the next section we inspect previous work on search interfaces on touch-enabled devices.

Search with Touch Devices

The workshop on exploratory search and Human-Computer Interaction at the 2007 CHI conference [37] demonstrates the interest in the research community on extending search interfaces to new kind of interactive environments. One of the discussed topics, in fact, was about the need to better understand how to design exploratory search systems for beyond-the-desktop interaction. There is a raising need for search system designs on interactive displays that take advantage of the idiosyncrasies of multi-touch interactions [39] instead of simply being directly ported from desktop-based interfaces.

One of the critical issues is to reduce the need for text entry, which is not suitable for touch-based interaction [42]. In the case of a small form-factor, FaThumb [18] explores web browsing on mobile phones by providing an interface that exploits facet navigation and limits text entry only to further narrow search results. Findings from the user evaluation demonstrated that text entry is more efficient for direct search while for open-ended search facet navigation offers better performances. Multi-touch gesture-based interaction has also been exploited as a mean to improve targeted search of specific content. For instance, Gesture Search [20] is a tool that allows users to define personalized touch gestures to quickly access data items on their mobile phone. The Questions not Answers (QnA) [16] prototype is an interesting instance of a system that exploits social interactions and context-awareness capabilities of mobile devices in order to offer reuse of previous queries based on geographical locations. The system maps other users' queries to their physical location and provides an interface to display them on an interactive map. Results from the user study demonstrated that displaying previous query result in less need to formulate new query or to a better formulation, which is influenced by the displayed queries.

Concerning large surfaces, like in public display settings, surface computing has been examined particularly for open-ended information exploration [14, 9]. Especially interesting cases are the EMDialog [13], which provides a visual exploration environment for an artist's work in a museum via temporal and contextual dimensions, and the Bohemian Bookshelf [33], which is designed to systematically support serendipitous discoveries while searching in a book collection through different interlinked visualisations. Insights from these research works led to several design principles for search interfaces in public spaces, such as combining different search strategies, rewarding short-term and long-term exploration and making information exploration appealing through engaging multi-touch interaction and information visualizations. Research literature emphasized on the use of large interactive surfaces for collaborative search tasks. Morris et al. [24] have conducted extensive empirical research on collaborative information seeking using horizontal surfaces, providing discussions on opportunities, challenges and design principles for the development of co-located search systems on tabletops along different dimensions such as collaboration styles, search input, group size and application domains. Efficient text entry by enabling the reuse of existing text instead of typing all searches on a virtual keyboard is also a leading design goal in the case of large surfaces [25].

From the state of the art, it is manifest that current solutions are modelling search systems for small or large surface and there is a lack of investigation on touch interfaces for medium-sized display screens, such as tablets. They present different affordances if compared to smaller or larger form-factor, and therefore need a different design approach. For instance, they do not support collaborative tasks as large surfaces do —given the limited screen size— but the display dimension is still bigger than mobile phones —while supporting users mobility— thus allowing richer visualisations, arrangements of interface elements and touch-based manipulations (e.g., two-handed gestures).

DESIGN CHALLENGES

From the state of the art, we identified three main challenges:

1. Formulating queries in unknown information spaces
Activities considered as related to exploratory search are very diverse and hard to define in a consolidated way. But unlike basic lookup search, they usually take place in areas that are unfamiliar to the user and are characterized by the frequent need to reformulate the query.

2. Learning about the information space and identifying possible search directions
The spatial metaphor of exploration applied to information retrieval well describes the need for steering the exploratory evolution in the information space. Narrowing possibilities to make steering decisions implies continuous gain of topical information.

3. Going through long lists of results with low information gain
In current search systems, users are forced to invest significant cognitive efforts in acquiring cues to formulate queries from the intermediate results, instead of focusing on collecting and learning from relevant information. Long lists of results conflict with the idea of a dynamic steering by slowing down the iterative process of query reformulation.

4. Typing and manipulating queries
Exploratory search activities performed on a traditional typed-query search interface require a lot of text entry and text manipulation. On touch devices, text manipulation is made difficult by the absence of a physical keyboard, hotkeys or shortcuts, and the lack of an accurate selection tool.

GENERAL DESIGN PRINCIPLES

From the above mentioned challenges, we came up with targeted design principles to guide the implementation of our prototype:

A. Flexible reuse and combination of information items to facilitate query formulation
To reduce the need for text entry (challenge 4), we chose to itemize information into entities of different types that can be flexibly manipulated and "dragged around" to support and facilitate all fundamental tasks like selection, duplication, grouping, deletion. Entities would be used to formulate queries, either individually or combined, to get a set of new entities as search results. An existing query could then be easily refined or reformulated by addition or removal of such entities and the results would update accordingly.

The possibility to input text is still necessary in some situations, for example when the system fails to make the proper suggestions or when specifying a first query. We decided our design should thus support it as an alternative and as a way to instantiate a search session.

B. Result sets of not only documents but most relevant entities to foster iterative query reformulation
To foster iterative query reformulation (challenge 3), we introduce the notion of *search streams* which describes an interactive structure supporting a query and related results: the query itself is formed of one or more entities and is composed by the user, while the results are shown as a vertical arrangement of entities related to the query and positioned above it. In the query area, items can be moved freely. Under a certain horizontal distance threshold, those entities are considered as a single query. The unity of a query is visualized through a network of thin lines linking the entities together. At first the query visually leads to a button that triggers the retrieval. The search engine then returns a set of entities related to the query. Those represents not only retrieved documents but also new entities, such as keywords or persons. They are vertically ordered by type and relevance. The flexibility of the search stream comes from its two level structure. It acts partly as a consolidated unit which can be moved around and considered as an almost traditional list of results, but each document or

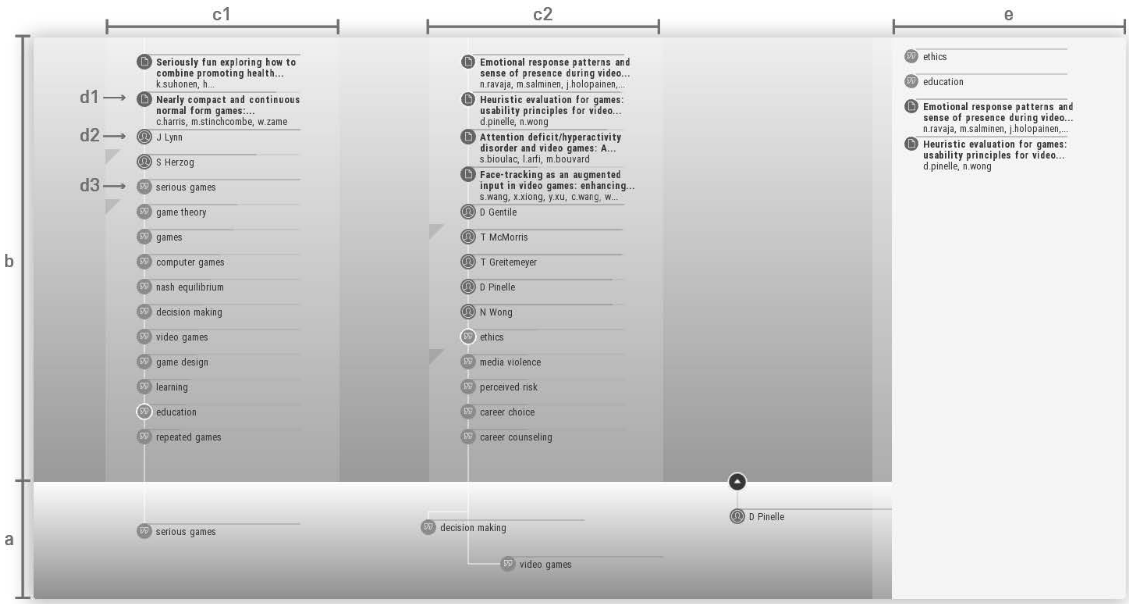


Figure 1. The Exploration Wall interface is composed of the (a) query area, (b) the results area, (c1, c2) search streams, three types of entities: (d1) documents (brown icon), (d2) authors (red icon), and (d3) keywords (blue icon), and (e) the reading-list drawer. The user can move any information item to compose queries by spatially grouping them on the query area. The whole workspace is scrollable and unlimited. Multi-touch gestures allows user to easily add or remove space between streams, or combine streams.

entity can become a new query, or part of an existing query, in the same stream or a parallel one.

C. Use of spatial configuration of multiple streams to identify search directions and learn about the information space

To facilitate steering decisions (challenge 2) and help the user formulate queries (challenge 1), our design supports search on simultaneous parallel streams. Persistency of search and context improves exploration by fostering trials without fear of losing current work, and supporting information comparison and entity association leading to quick instantiation of new queries or quick query re-formulation. It also allows the user to keep track of former queries and results while supporting unconstrained branching and revisits in the actual search process.

EXPLORATION WALL

Here we describe the interactions and implementation of the system based on the above mentioned design principles.

The User Interface

The interface of Exploration Wall is entirely dedicated to its main workspace (Figure 1), which is divided in two areas: the query area at the bottom (Figure 1-a) and the result area on top (Figure 1-b). The workspace supports information in the form of parallel search streams (Figure 1-c1 and c2) organized by taking advantage of the multi-touch ability: it can be scrolled on the horizontal axis with a simple swipe gesture on the background, horizontal space can be added or removed at will from a specific location using a conventional pinch gesture, the same pinch gesture can also be used to dilate or contract space (e.g., to quickly improve legibility of an area cramped with information).

In current instantiation, entities are of three types (Figure 1): Documents (Figure 1-d1), Authors (Figure 1-d2) and Keywords (Figure 1-d3). Each entity is represented by a pictogram, a label and a relevance gauge. One can move an entity by dragging its pictogram. Additional interactions include: tap on the title of a document to reveal additional information like source and content, tap on the icon to store the entity. Stored entities appear highlighted and can be found in the storage drawer described below.

The storage drawer (Figure 1-e) offers an unobtrusive solution that acts as a reading list as well as an always accessible storage area for information transit. One opens and closes it by performing a swipe gesture from the right edge of the display.

The Search Engine

The search engine was designed to support multi-touch interaction design of Exploration Wall and is based on two design rationale. First, the *entity ranking* where entities that are returned for the user to manipulate and use to formulate queries should be as central to the topic as possible. For example, if the user searches for "information retrieval", she is not expecting back only entities that occur in the top ranked

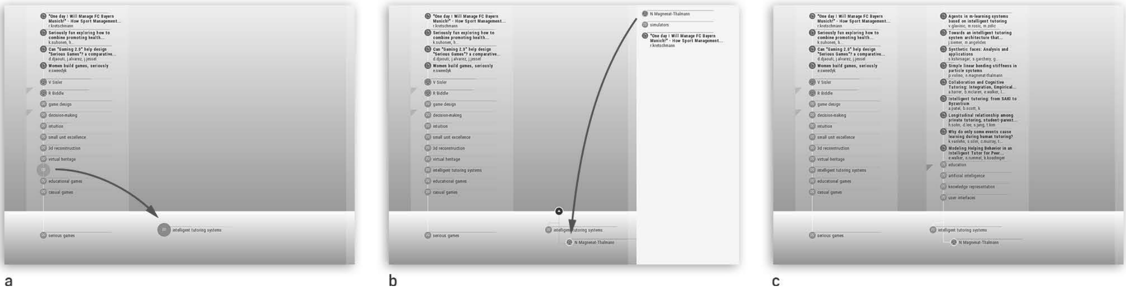


Figure 2. An exemplary search scenario illustrating the functionality of the Exploration Wall system. (a) After initiating a query the user receives a set of results, the user notices a keyword-entity that tackles her interest and drags it to the query area. (b) The user investigates the selected keyword in relation to an author entity that has been formerly saved in to the reading list by dragging the author entity to the query area as well, close enough so they become visually associated. (c) The user taps on the trigger to retrieve a new set of documents and entities (authors and keywords), that can be further manipulated and used to combine with an existing search stream or to create a new search stream.

documents, but that are central for the field of information retrieval. Second, the *document ranking* where the documents that are returned for the user as results after making some query, say "information retrieval" and "relevance feedback" should be not the most central entities, but the most relevant documents matching the query.

Entity Ranking

We represent the data as an undirected graph, where each document, keyword, and author are represented as vertices and the edges represent their occurrence in the document data.

The centrality ranking is based on the user's relevance feedback on vertices determined by dragging them into the query area. Each cluster in a query area represents a separate query that consists of a set of vertices. We use the personalized PageRank method [15] to compute the ranking of the vertices. The set of nodes that the user has chosen to be part of an individual query form the personalization vector that is set to be the prior for the PageRank computation [15]. We compute the steady distribution by using the power iteration method with 50 iterations. The top k=10 nodes from each entity category (keyword, author) are selected for presentation for the user.

Document Ranking

The document ranking is based on language modelling approach of information retrieval [41], where a unigram language model is built for each document and the maximum likelihood of the document generating the query is used to compute the ranking. We use Jelinek-Mercer smoothing to avoid zero probabilities in the estimation.

Intuitively, separating the entity ranking and document ranking approaches makes it possible to compute a limited set of entities that are likely to be the most important in the graph given the user interactions and allows users to target their feedback on a subset of the most central nodes given the interaction history of the user in any subsequent iteration. At the same time, the document ranking enables accurate and well-established methodology for ensuring relevance of the documents.

EVALUATION

The main purpose of the evaluation was to observe the effects and implications of the design of Exploration Wall on search performance and search behavior. Therefore, Exploration Wall was compared to a conventional search interface which was used as a baseline. The experiment concerned the following factors: effectiveness, expert rating, search behavior, usability and user engagement. The evaluation was composed of two tasks, a short one (5 minutes) and a long one (20 minutes).

Dataset

We used a document set including over 50 million scientific documents from the following data sources: the Web of Science prepared by Thomson Reuters, Inc., the Digital Library of the Association of Computing Machinery (ACM), the Digital Library of Institute of Electrical and Electronics Engineers (IEEE), and the Digital Library of Springer. The information about each document consists of: title, abstract, author names, and publication venue. Both the baseline and Exploration Wall used the same document set.

We decided to limit the data to scientific literature for two reasons. First, the data should allow retrieval tasks that result in exploration, and scientific search tasks are suitable for scenarios where users' goals are uncertain and require exploratory search behavior. Second, experts were available for providing high quality relevance assessments for task outcomes.

Baseline

The baseline, shown on Figure 3, was implemented following the interface principles of traditional search tools: typed query and resulting list of returned documents presented by title, with authors and keywords. The system uses the same dataset used by Exploration Wall to permit comparability. Also, the ranking is based on the same document retrieval model as in Exploration Wall, but to mimic traditional search engines it ranks only documents, while authors and keywords are only shown as additional information associated to each document. Last, our system did not allow dynamic updates of the search result when typing the query. All these factors aimed to create a baseline allowing us to focus the evaluation solely on the user interface design of Exploration Wall.

Tasks

The evaluation was composed of two tasks, a short one and a long one. We chose 6 possible different topics for the two tasks: *crowdsourcing*, *smartphones energy efficiency*, *diagrams*, *semantic web*, *lie detection* and *digital audio effects*. In order to ensure that participants were not experts in the topics and could perform a real exploratory search, they prerated their familiarity with the topics on a 1 (less familiar) to 5 (most familiar) scale. The four less familiar topics were used in the tasks. Both tasks were performed with different topics, so the participants did not the know the results from the previous task.

Short Task

For this task, we asked the users: "Search and list 5 relevant authors, documents and keywords that you consider relevant in topic Y." The time limit for this task was 5 minutes.

Long Task

For this task, we asked the users: "Imagine that you are writing a scientific essay on the topic X. Search and collect as many relevant scientific documents as possible that you find useful for this essay. During the task, please, list what you think are the top five key technologies, persons, documents and research areas and write five bullet lines, which would work as the core content of the essay." The time limit for this task was 20 minutes.

Participants and Procedure

We recruited 10 researchers from the computer science departments of two universities with a range of research experience. The 20% of them were females, which matched the gender ratio of both departments, and the mean age was M=30.5, SD=5.52. For the experiment participants used an iPad Air Wi-Fi tablet, as shown on Figure 4.

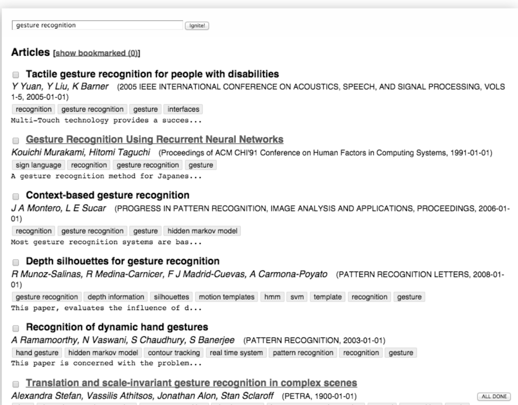


Figure 3. A screenshot of the baseline system that uses the same underlying document set and ranking model, and allows typed-query interaction.

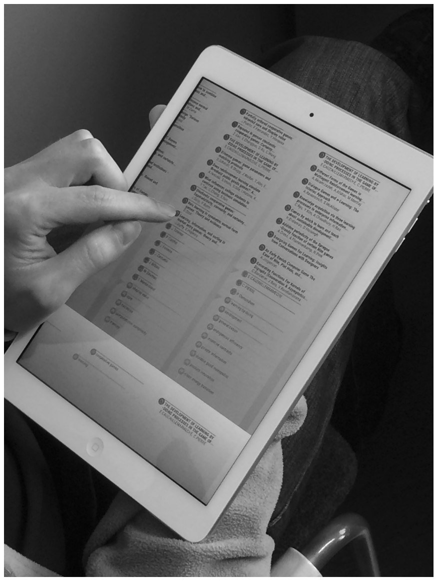


Figure 4. Exploration Wall was evaluated using using the iPad Air Wi-Fi tablet.

In this study, we followed a within-subjects experiment design, counter-balanced by changing the order of the two tested interfaces, as well as the order of the two tasks. Before starting the main tasks, users received detailed instructions on how to use the interface and performed a 5 minutes training task on each interface. For text entry, we relied on the native virtual keyboard of the tablet. At the end of the sessions participants were asked to answer the UES and SUS questionnaires for each interface via on-line forms (Google Forms). We used the API and service of logentries.com to log all actions and data.

MEASURES

The experiment considered the following factors: effectiveness, expert rating, search behavior, usability and user engagement which were measured as follows.

Effectiveness

The effectiveness refers to the quality of the information retrieved and displayed by a system. Since our baseline system returns lists of documents while Exploration Wall returns lists of mixed-type entities, we chose to solely measure the quality of the displayed documents. We created ground truth by pooling the retrieved documents from the system logs. Domain experts were then asked to assess the relevance of the retrieved documents on a binary scale (relevant or irrelevant). Effectiveness was measured by precision, recall and F-measure at two levels [21]. First, we measured the average retrieval effectiveness at a query level as an average quality of the documents returned in response to a user interaction. Second, we measured the retrieval effectiveness at task level as an cumulative quality of documents retrieved within the whole search session.

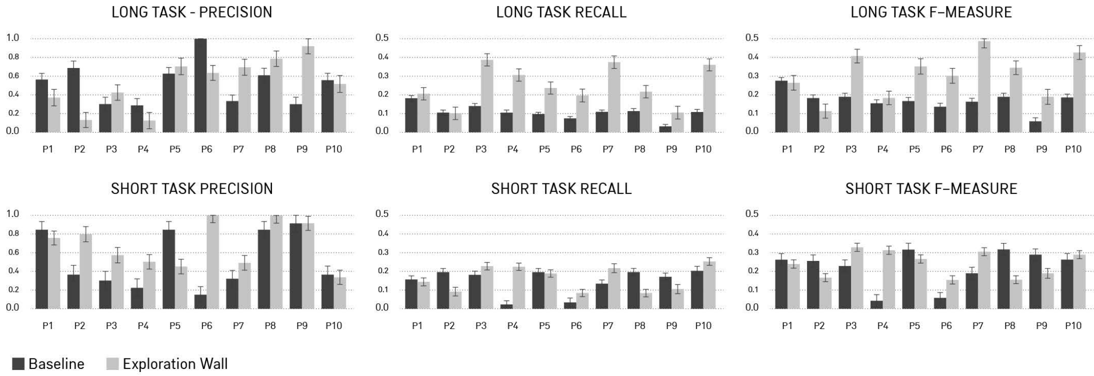


Figure 5. Effectiveness results for the short and long tasks split by participants. Results are reported as the mean of every query-response of each participant during the task.

Expert Rating

Experts were asked to rate the task outcome. For the short task, the outcome was a list of documents, and two types of entities: authors and keywords. The relevance of each item was evaluated on a 5-point scale (1 less relevant - 5 most relevant). The outcome of the long task was an essay, a set of documents, and a set of entities: keywords representing technologies and research areas, and persons. The sets of documents and entities were evaluated in the same way as in the short task, while the essay was evaluated on a different 5-point scale (5=Excellent, 4=Good, 3=Satisfactory, 2=Deficient, 1=Failing).

Search Trail Analysis

In order to understand and compare users' search behavior, we logged user actions and extracted corresponding search trails using a method resembling White's [36]. In a similar manner, we then looked for descriptive statistics of the search trails by selecting six parameters relevant to both interfaces.

- Number of queries: the total number of queries that were submitted during each task on both interface.
- Number of text entries per query
- Number of revisits: The number of revisits to a query or stream consulted earlier in the current trail.
- Number of branches: The number of times a subject revisited a query or stream on the current trail and then proceeded with formulation of a new query.
- Number of queries/min: the number of queries per minute that were submitted during each task on both interface.
- Number of parallel queries: Number of parallel streams produced with Exploration Wall or number of tabs opened with the baseline.

Usability and Engagement

As usability assessment questionnaires we used the standard System Usability Scale (SUS) [6] and the User Engagement Scale (UES) for exploratory search [26]. SUS consists of a ten item questionnaire and is a widely used and validated for measuring perceptions of usability. Since the degree of user engagement is a strong indicator of exploratory search performance [38], we chose to use UES for exploratory search. The User Engagement Scale (UES) questionnaire include 27 questions considering six different dimensions: Aesthetics (AE), Focused Attention (FA), Felt Involvement (FI), Perceived Usability (PUs), Novelty (NO) and Endurability (EN) aspects of the experience.

RESULTS

In this section, we present results from the user experiments divided according to the different factors: effectiveness, expert rating, search trail analysis, and usability and engagement.

Effectiveness

The effectiveness results are given in Table 1. The results show that Exploration Wall shows substantial improvement in the long task. The improvement was found to hold for task-level measurement, but also for averaged interaction-level measurement for which the recall and the F-measure were found to be significantly higher compared to the baseline. On average at the query level, the F-measure for the Exploration Wall was improved (M=0.136, SD=0.122). This im-

	Long Task			Short Task		
	BL	EW	p	BL	EW	p
P (Task)	0.40	0.42	0.85	0.52	0.58	0.67
R (Task)	0.13	**0.38**	**<0.01**	0.18	0.21	0.59
F (Task)	0.17	**0.34**	**<0.01**	0.25	0.26	0.90
P (Query)	0.53	0.53	0.96	0.52	0.69	0.16
R (Query)	0.11	**0.25**	**<0.01**	0.15	0.16	0.69
F (Query)	0.17	**0.31**	**<0.01**	0.22	0.24	0.41

Table 1. Effectiveness results for the short and long tasks. Results are reported cumulatively for the whole duration of the task and as a mean of every query-response during the task. P=Precision, R=Recall, F=F_1 measure, EW=Exploration Wall, BL=Baseline.

	Long Task						
Search Trail Features	BL			EW			BL vs EW
	M	SD	Median	M	SD	Median	Wilcoxon Test
No. of queries	4.30	3.09	4.50	12.10	6.97	**13.50**	**Z = -2.76, p <0.01**
No. of text entries/query	1.00	0.00	1.00	0.36	0.35	**0.27**	**Z = 2.67, p <0.01**
No. of branches	0.10	0.31	0.00	5.70	4.55	**6.00**	**Z = -2.68, p <0.01**
No. of revisits	0.70	1.64	0.00	7.00	6.09	**6.00**	**Z = -2.67, p <0.01**
No. of queries/min	0.26	0.17	0.26	0.63	0.36	**0.70**	**Z = -2.70, p <0.01**
No. parallel queries	1.70	1.06	1.00	8.50	5.89	**7.00**	**Z = -2.76, p <0.01**
	Short Task						
Search Trail Feature	BL			EW			BL vs EW
	M	SD	Median	M	SD	Median	Wilcoxon Test
No. of queries	2.50	1.58	2.00	3.50	2.12	4.00	Z = -1.46, p >0.05
No. of text entries/query	1.00	0.00	1.00	0.55	0.35	**0.47**	**Z = 2.55, p <0.05**
No. of branches	0.00	0.00	0.00	0.8	1.03	**0.5**	**Z = -2.21, p <0.05**
No. of revisits	0.20	0.42	0.00	1.1	1.10	1.0	Z = -1.81, p >0.05
No. of queries/min	0.59	0.33	0.45	0.86	0.36	0.93	Z = -2.24, p >0.05
No. of parallel queries	1.30	0.67	1.00	2.70	2.00	**2.00**	**Z = -2.40, p <0.05**

Table 2. Results of the search trail analysis for the short and long tasks. Means, Standard Deviation, Median (used in the Wilcoxon Matched-Pairs test) as well as Significant differences of search trail feature considering both interfaces. The values in bold show the significant differences. BL=baseline, EW=Exploration Wall.

provement was statistically significant, $t(9)$=3.519, $p < 0.01$. This is a direct consequence of the improvement in the recall (M=0.142, SD=0.094, $t(9)$=4.790, $p < 0.001$). The difference in precision was not significant (M=0.005, SD=0.366) which indicates that while Exploration Wall improves recall it retains precision. In terms of effectiveness, no statistically significant differences between the systems were found in the short task.

Figure 5 shows the query-level effectiveness for the long tasks and the short task split by participants. Exploration Wall constantly outperforms the baseline system in terms of recall and F-measure in the long task. The effect is steady across participants. No significant differences between the systems were found in the short task.

Expert Rating

Unlike the effectiveness, the expert rating showed no significant differences between the Exploration Wall and the Baseline. Regarding the relevance of selected items, the mean values for the long task were M=3.54, SD=0.67 for Exploration Wall and M=3.45, SD=0.82 for Baseline, while for short task they were M=3.60, SD=1.23 for Exploration Wall and M=3.83, SD=0.99 for Baseline. Regarding the the relevance of the essays produced in the end of the long task the mean values were M=3.90, SD=0.75 for Exploration Wall and M=4.05, SD=0.69 for Baseline.

Search Trail Analysis

Table 2 shows the results of the search trail analysis. The Shapiro-Wilk test indicated that the search trail data did not follow a normal distribution, and the Wilcoxon Matched-Pairs test was used for significance testing. The users in the Exploration Wall condition were found to use all of the measured interaction features significantly more than the users in the baseline condition in the long task. Differences were also found in the short task. The users in the Exploration Wall condition typed less, branched more, and used more parallel queries.

Usability and Engagement

The results for the mean of answers of the SUS questionnaire, i.e., for usability, were M=78.85, SD=12.43 for Exploration Wall and M=62.25, SD=15.65 for the baseline. A paired t-test showed a significant difference ($t(9)$=2.36, $p < 0.05$) between the two systems, revealing higher usability for Exploration Wall. The results of the UES questionnaires are also favorable for Exploration Wall. Wilcoxon Matched-Pairs test shows that in 70% of the questions there is a significant difference between the interfaces, all in favour of Exploration Wall.

DISCUSSION AND CONCLUSIONS

Challenges in supporting exploratory search include providing resources for formulating queries in unknown areas [11], learning about possible directions in the information space [29], and going through long list of results with low information gain [38, 5]. In particular on keyboard-less touch devices the challenges are aggravated by typing efforts. We introduced Exploration Wall a novel user interface that addresses these challenges with a principled design. The founding principle is to transform results into entities that can be flexibly manipulated and used for creating queries and search streams. The wall is a canvas where parallel and previous search streams are juxtaposed and provide a spatial exploration of the information space and possible exploration directions. The manipulation includes inspecting relationship between entities and facilitating the creation of new search streams.

The study shows how Exploration Wall is an effective tool for exploratory search on touch surfaces. Participants using Exploration Wall were able to exploit parallel search streams to iteratively refine their queries and deeply explore the search tree. The difference in recall proves that more relevant documents were retrieved when using Exploration Wall.

Exploration Wall also led to a more active search behavior, with more queries per minute and more branches. In addition, if we consider the fact that participants used more parallel queries with Exploration Wall (parallel streams) than with the baseline (parallel tabs), we can conclude that the participants took advantage of parallel streams with consequent avoidance of text input.

Results from the UES questionnaire also show a better user engagement, a factor that is likely to have contributed to the more active search behavior. In addition, the SUS scale shows that Exploration Wall presents a better usability than conventional search interfaces on tablets.

The study confirms how our design approach facilitates query formulation, by directing exploration in unknown areas, and providing alternatives to text inputs. While little or no differences were appreciated in short tasks, Exploration Wall proved to be an effective tool for long tasks by showing improved recall while preserving precision, as well as improved user engagement and satisfaction.

In addition to the positive results, this work is adaptable to many applications and setups that would enable new possibilities to be found through deeper study of user behavior (e.g. search strategies and nature of composed queries). It has important implications for future development of exploratory search systems in particular considering multimodal interaction and user interface for entity oriented search [27]. The principles are applicable to other datasets such as for example news search [4] as well as other devices and sizes (e.g. large multi-touch screen for collaborative work, mobile devices for mobility and privacy, combinations of devices, desktop).

Considering this, as well as the growing popularity of touch devices, our work offers a powerful and flexible template to be considered when designing user interfaces supporting exploratory search.

ACKNOWLEDGMENTS

This research was partially funded by TEKES (through the Re:KnoW project), the European Commission through the FP7 Project MindSee 611570, the Academy of Finland (278090), TIPEx Project (TIN2010-19859-C03-01) and "Postdoc Mobility Scholarship Programme" of Universidad Carlos III de Madrid. Certain data included here are derived from the Web of Science prepared by Thomson Reuters, Inc., Philadelphia, PA. Data are also included from the Digital Libraries of the ACM, IEEE, and Springer.

REFERENCES

1. Ahlberg, C., and Shneiderman, B. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. CHI*, ACM (New York, NY, USA, 1994), 313–317.
2. Ahn, J., and Brusilovsky, P. Adaptive visualization for exploratory information retrieval. *Information Processing & Management 49*, 5 (2013), 1139–1164.
3. Ahn, J.-w., and Brusilovsky, P. Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization 8*, 3 (2009), 167–179.
4. Andolina, S., Klouche, K., Peltonen, J., Hoque, M., Ruotsalo, T., Cabral, D., Klami, A., Głowacka, D., Floréen, P., and Jacucci, G. Intentstreams: smart parallel search streams for branching exploratory search. In *Proc. IUI*, ACM (2015).
5. Athukorala, K., Oulasvirta, A., Głowacka, D., Vreeken, J., and Jacucci, G. Narrow or broad?: Estimating subjective specificity in exploratory search. In *Proc. CIKM*, ACM (New York, NY, USA, 2014), 819–828.
6. Brooke, J. Sus-a quick and dirty usability scale. *Usability evaluation in industry 189* (1996), 194.
7. Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., and Zhang, Y. Effects of structure and interaction style on distinct search tasks. In *Proc. JCDL*, ACM (2007), 442–451.
8. Chau, D. H., Kittur, A., Hong, J. I., and Faloutsos, C. Apolo: Making sense of large network data by combining rich user interaction and machine learning. In *Proc. CHI*, ACM, ACM (New York, NY, USA, 2011), 167–176.
9. Coutrix, C., Kuikkaniemi, K., Kurvinen, E., Jacucci, G., Avdouevski, I., and Mäkelä, R. Fizzyvis: designing for playful information browsing on a multitouch public display. In *Proc. DPPI*, ACM (2011), 27.
10. Esenther, A., and Ryall, K. Fluid dtmouse: Better mouse support for touch-based interactions. In *Proc. AVI*, ACM (New York, NY, USA, 2006), 112–115.
11. Głowacka, D., Ruotsalo, T., Konyushkova, K., Athukorala, K., Jacucci, G., and Kaski, S. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proc. IUI*, ACM (New York, NY, USA, 2013), 117–128.
12. Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. Interactive visualization of multiple query results. In *Proc. InfoVis*, IEEE Computer Society (Washington, DC, USA, 2001), 105.
13. Hinrichs, U., Schmidt, H., and Carpendale, S. Emdialog: Bringing information visualization into the museum. *Visualization and Computer Graphics, IEEE Transactions on 14*, 6 (2008), 1181–1188.
14. Jacucci, G., Morrison, A., Richard, G. T., Kleimola, J., Peltonen, P., Parisi, L., and Laitinen, T. Worlds of information: designing for engagement at a public multi-touch display. In *Proc. CHI*, ACM (2010), 2267–2276.
15. Jeh, G., and Widom, J. Scaling personalized web search. In *Proc. WWW*, ACM (New York, NY, USA, 2003), 271–279.

16. Jones, M., Buchanan, G., Harper, R., and Xech, P.-L. Questions not answers: A novel mobile search technique. In *Proc. CHI*, ACM (New York, NY, USA, 2007), 155–158.

17. Käki, M. Findex: Search result categories help users when document ranking fails. In *Proc. CHI*, ACM (New York, NY, USA, 2005), 131–140.

18. Karlson, A. K., Robertson, G. G., Robbins, D. C., Czerwinski, M. P., and Smith, G. R. Fathumb: A facet-based interface for mobile search. In *Proc. CHI*, ACM (New York, NY, USA, 2006), 711–720.

19. Kules, W., Wilson, M. L., Schraefel, M. C., and Shneiderman, B. From keyword search to exploration: How result visualization aids discovery on the web. Tech. rep., University of Southampton, 2008.

20. Li, Y. Gesture search: a tool for fast mobile data access. In *Proc. UIST*, ACM (2010), 87–96.

21. Manning, C. D., Raghavan, P., and Schütze, H. *Introduction to information retrieval*, vol. 1. Cambridge university press Cambridge, 2008.

22. Marchionini, G. Exploratory search: from finding to understanding. *Commun. ACM 49*, 4 (Apr. 2006), 41–46.

23. Matejka, J., Grossman, T., and Fitzmaurice, G. Citeology: Visualizing paper genealogy. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, ACM (New York, NY, USA, 2012), 181–190.

24. Morris, M. R., Fisher, D., and Wigdor, D. Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. *Information processing & management 46*, 6 (2010), 703–717.

25. Morris, M. R., Lombardo, J., and Wigdor, D. Wesearch: supporting collaborative search and sensemaking on a tabletop display. In *Proc. CSCW*, ACM (2010), 401–410.

26. O'Brien, H. L., and Toms, E. G. Examining the generalizability of the user engagement scale (ues) in exploratory search. *Information Processing & Management 49*, 5 (2013), 1092 – 1107.

27. Ruotsalo, T., and Hyvönen, E. Exploiting semantic annotations for domain-specific entity search. In *Proc. ECIR*, Springer (2015).

28. Ruotsalo, T., Jacucci, G., Myllymäki, P., and Kaski, S. Interactive intent modeling: Information discovery beyond search. *Communications of the ACM 58*, 1 (Jan. 2015), 86–92.

29. Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K., Kosunen, I., Reijonen, A., Myllymäki, P., Jacucci, G., et al. Directing exploratory search with interactive intent modeling. In *Proc. CIKM*, ACM (2013), 1759–1764.

30. Sadler, E. B. Project blacklight: a next generation library catalog at a first generation university. *Library Hi Tech 27*, 1 (2009), 57–67.

31. Stasko, J., Görg, C., and Liu, Z. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization 7*, 2 (2008), 118–132.

32. Terveen, L., Hill, W., and Amento, B. Constructing, organizing, and visualizing collections of topically related web resources. *ACM Transactions on Computer-Human Interaction 6*, 1 (1999), 67–94.

33. Thudt, A., Hinrichs, U., and Carpendale, S. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proc. CHI*, ACM (2012), 1461–1470.

34. Varcholik, P. D., LaViola, J. J., and Hughes, C. E. Establishing a baseline for text entry for a multi-touch virtual keyboard. *Int. J. of Human-Computer Studies 70*, 10 (2012), 657 – 672.

35. Verbert, K., Parra, D., Brusilovsky, P., and Duval, E. Visualizing recommendations to support exploration, transparency and controllability. In *Proc. IUI*, ACM, ACM (New York, NY, USA, 2013), 351–362.

36. White, R. W., and Drucker, S. M. Investigating behavioral variability in web search. In *Proc. WWW*, ACM (2007), 21–30.

37. White, R. W., Drucker, S. M., Marchionini, G., Hearst, M., and schraefel, m. c. Exploratory search and hci: Designing and evaluating interfaces to support exploratory search interaction. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, ACM (New York, NY, USA, 2007), 2877–2880.

38. White, R. W., and Roth, R. A. Exploratory search: Beyond the query-response paradigm. *Synth. Lec. on Inf. Conc., Retr., and Serv. 1*, 1 (2009), 1–98.

39. Wigdor, D., and Wixon, D. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011.

40. Wildemuth, B. M., and Freund, L. Assigning search tasks designed to elicit exploratory search behaviors. In *Proc. HCIR*, ACM (New York, NY, USA, 2012), 4:1–4:10.

41. Zhai, C., and Lafferty, J. A study of smoothing methods for language models applied to information retrieval. *ACM Trans. Inf. Syst. 22*, 2 (2004), 179–214.

42. Zhai, S., Kristensson, P.-O., and Smith, B. A. In search of effective text input interfaces for off the desktop computing. *Interacting with computers 17*, 3 (2005), 229–250.

Publication II

Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, (IUI '15). ACM, New York, NY, USA, pp. 300-305, 2015. DOI: 10.1145/2678025.2701401, June 2015.

IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search

Salvatore Andolina[1], **Khalil Klouche**[1,4], **Jaakko Peltonen**[2,3], **Mohammad Hoque**[1], **Tuukka Ruotsalo**[2], **Diogo Cabral**[1], **Arto Klami**[1], **Dorota Głowacka**[1], **Patrik Floréen**[1,2], **and Giulio Jacucci**[1,2]
Helsinki Institute for Information Technology HIIT
[1]University of Helsinki, Department of Computer Science, Finland
[2]Aalto University, Finland [3]University of Tampere, School of Information Sciences, Finland
[4]Aalto University, Media Lab, Finland

ABSTRACT

The user's understanding of information needs and the information available in the data collection can evolve during an exploratory search session. Search systems tailored for well-defined narrow search tasks may be suboptimal for exploratory search where the user can sequentially refine the expressions of her information needs and explore alternative search directions. A major challenge for exploratory search systems design is how to support such behavior and expose the user to relevant yet novel information that can be difficult to discover by using conventional query formulation techniques. We introduce IntentStreams, a system for exploratory search that provides interactive query refinement mechanisms and parallel visualization of search streams. The system models each search stream via an intent model allowing rapid user feedback. The user interface allows swift initiation of alternative and parallel search streams by direct manipulation that does not require typing. A study with 13 participants shows that IntentStreams allow better support for branching behavior compared to a conventional search system.

Author Keywords

User interface design; information exploration; parallel browsing.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

Exploratory search activities confront users with problems in formulating queries and identifying directions for information exploration. Studies show that searchers tend to perform more than one task simultaneously: approximately 75% of submitted queries involve a multitasking activity [22, 21]. Users engage in multitask search with and without parallel browsing, but parallel browsing is a common activity and more prevalent than linear browsing. In parallel browsing, also called branching [13], users visit web pages in multiple concurrent threads, for example, by opening multiple tabs or windows in web browsers [14]. Branching in browsing has been studied extensively [6], but little has been done to support nonlinear and parallel browsing. Recent visual search user interfaces have shown the effectiveness of interacting visually with query elements, however, there are no solutions to support fluid branching and parallel search.

We introduce IntentStreams, a system supporting parallel browsing and branching during search without the need to open new tabs. It presents parallel streams of searches, where each stream shows a list of resulting documents and keywords, and a display of the underlying queries as keywords representing the search intent of the stream. New streams are initiated by the user, where the search intent of a new stream is initialized by typing a traditional query or by dragging keywords available in any of the streams. In each stream, in addition to the user-chosen keywords, the system proposes other relevant keywords and orders them vertically by their predicted relevance. The users can change the relative relevance of keywords in the query intent of each stream and branch new streams by simply dragging keywords. IntentStreams was tested using 25 million news articles crawled from public news sources in a comparative study with 13 subjects.

The experimental results show that IntentStreams better supports parallel search and branching behavior when compared to a conventional search system.

Exploratory search is commonly distinguished from lookup search and has been described as combining exploratory browsing and focused searching [23, 5]. These and other information seeking frameworks [23, 27] partially overlap, emphasizing different aspects but similarly trying to characterise iterative sense making and refinement of search intents. As a response, personalization techniques have been developed to support query formulation and relevance feedback. Several techniques exist for supporting query formulation or for processing results to help re-rank, filter [31] or expand the query [8, 26]. For example, relevance feedback [16] and query or term suggestions [17] can be effective in short-term navigational search, but give limited support for exploratory

IUI 2015, March 29–April 1, 2015, Atlanta, GA, USA.

http://dx.doi.org/10.1145/2678025.2701401

search. However, increasing evidence in information retrieval research suggests relevance feedback in exploratory search often leads to a context trap where after a few iterations of feedback users have specified their context so strictly the system is unable to propose anything new [12, 23].

Visual search interfaces. Recently, search systems use visualization of the resulting information for faster relevance judgment and effective feedback [11, 15, 20, 24, 32]. Several approaches to visualize results have been explored, including multiple linked lists, scatter plots, graphs and their combinations [30, 20]. Such visual search systems are distinguished from traditional query composition ones through their emphasis on rapid filtering to reduce result sets, progressive refinement of search parameters, continuous reformulation of goals, and visual scanning to identify results [1]. Visual approaches provide better support for exploration in a variety of ways: supporting sense making by incrementally and interactively exploring the network of data [7], showing how visualization supports user involvement in the recommendation providing rationale behind suggested items [34], visualizing relations of different queries and result sets [2], supporting entity search on touch devices [18]. Recent work shows how users can view and manipulate their search intent models provided by the system [2, 3, 29, 9, 28]. However, there is still a lack of solutions that specifically address branching behavior while supporting query formulation.

INTENTSTREAMS

A user performing exploratory search typically needs to steer the exploration by query refining or reformulating as she makes better sense of the information space. We identified two challenges impairing a proper steering support:

1. Supporting information comparison from several queries: Traditional search systems show one set of results at a time. To compare information from distinct queries, one must rely on memory or complicated browser windows arrangements.

2. Formulating queries in unknown areas: Exploratory search tasks usually require the user to iteratively formulate queries in a field she is unfamiliar with. To progress with the search, the used needs to go through long lists of results with low information gain to acquire new terms and notions to reformulate the query.

We tackle these challenges through a combination of a user interface supporting parallel search streams, and user intent modeling.

The User Interface

IntentStreams provides a unique horizontally scrollable workspace divided in two areas: the keywords area at the bottom and the results area on top (Figure 1). By clicking the workspace, the user is prompted to type a first query. The system returns a list of relevant documents in the results area and a set of related keywords in the keywords area. Keywords are positioned vertically by weight and horizontally by topic proximity. The vertical arrangement is called a stream and can be easily manipulated, modified and refreshed. The content of a document can be seen by clicking the title. A click and hold on a document highlights keywords directly related to it. A click and hold on a keyword highlights related documents. By moving keywords vertically, the user can change their weight; by hitting the refresh button, the stream then updates and presents a new set of documents and keywords. New parallel streams can be created by clicking next to an existing stream and typing a new query, or simply by dragging a keyword outside of its stream. Since the workspace is horizontally scrollable, the amount of parallel streams a user can create is limited only by computer memory. The amount of parallel streams that can be shown simultaneously is determined by the display resolution. Streams can be dragged and rearranged. A button lets the user delete streams.

Interactive Intent Model

For each search stream, the interactive intent model is similar to the model in a previous non-parallel system [29] and has two parts: a model for retrieval of documents, and a model for estimating the user's search intent (relevance of keywords to the user's information need). We describe both below.

Document retrieval model. For each stream, we estimate a relevance ranking where documents are ranked by their probability given the intent model for the stream. We use a unigram language model. The intent model yields a vector $\hat{\mathbf{v}}$ with a weight $\hat{v}_i$ for each keyword k_i. The $\hat{\mathbf{v}}$ is treated as a sample of a desired document. Documents d_j are ranked by probability to observe $\hat{\mathbf{v}}$ as a sample from the language model M_{d_j} of d_j. Maximum likelihood estimation yields $\hat{P}(\hat{\mathbf{v}}|M_{d_j}) = \prod_{i=1}^{|\hat{\mathbf{v}}|}(\hat{P}_{mle}(k_i|M_{d_j}))^{\hat{v}_i}$. We regularize probabilities $\hat{P}_{mle}(k_i|M_{d_j})$ in d_j towards overall keyword proportions in the corpus by Bayesian Dirichlet smoothing. In each stream the d_j are ranked by $\alpha_j = \hat{P}(\hat{\mathbf{v}}|M_{d_j})$. To expose the user to more novel documents we sample a document set from the ranking and show them in rank order. As in [9, 29] we use Dirichlet Sampling based on the α_j, and favor documents whose keywords got positive feedback by increasing their α_j [10].

User intent model. For each stream, the intent model estimates relevance of keywords from feedback to keywords. For a stream launched by a typed query, we use the query with weight 1 as the initial intent model; for a stream launched by dragging a keyword we use the keyword with weight 1. The user gives feedback as relevance scores $r_i \in [0, 1]$ for a subset of J keywords $k_i, i = 1, \ldots, J$ in the stream; $r_i = 1$ means k_i is highly relevant and the user wishes to direct the stream in that direction, and $r_i = 0$ means k_i is of no interest.

Let $\mathbf{k}_i$ be binary $n \times 1$ vectors telling which of the n documents k_i appeared in; to boost documents with rare keywords we convert the $\mathbf{k}_i$ to *tf-idf* representation. We estimate the expected relevance r_i of a keyword k_i as $\mathbb{E}[r_i] = \mathbf{k}_i^\top \mathbf{w}$. The vector $\mathbf{w}$ is estimated from user feedback by the LinRel algorithm [4]. In each search iteration, let $k_1, \ldots, k_p$ be the keywords for which the user gave feedback so far, let $\mathbf{K} = [\mathbf{k}_1, \ldots, \mathbf{k}_p]^\top$ be the matrix of their feature vectors, and let $\mathbf{r}^{feedback} = [r_1, r_2, \ldots, r_p]^\top$ be their relevance scores from the user. LinRel estimates $\hat{\mathbf{w}}$ by solving $\mathbf{r}^{feedback} = \mathbf{K}\mathbf{w}$, and estimates relevance score for each k_i as $\hat{r}_i = \mathbf{k}_i^\top \hat{\mathbf{w}}$.

Figure 1. a. The first query (in this case mobile phone") returns a search stream composed of news articles most relevant to the query, as well as a set of most relevant keywords extracted from a larger set of related articles. b. The user can modify the weight of the keywords by sliding them vertically, after which the stream will refresh, updating articles and keywords accordingly. If dropped outside their initial stream, keywords can either trigger a new search stream or be passed to an already existing parallel stream.

To expose the user to novel keywords, in each stream we show keywords k_i not with highest $\hat{r}_i$, but with highest upper confidence bound for relevance, which is $\hat{r}_i + \alpha\sigma_i$, where σ_i is an upper bound on standard deviation of $\hat{r}_i$, and $\alpha > 0$ is a constant for adjusting the confidence level. In each iteration, we compute $\mathbf{s}_i = \mathbf{K}(\mathbf{K}^\top\mathbf{K}+\lambda\mathbf{I})^{-1}\mathbf{k}_i$ where λ is a regularization parameter, and show the k_i maximizing $\mathbf{s}_i^\top \mathbf{r}^{feedback} + \frac{\alpha}{2}\|\mathbf{s}_i\|$ representing estimated search intent. We optimize horizontal positions of the shown k_i by dimensionality reduction [33]; k_i get similar positions if their relevance estimate changes similarly with respect to a set of additional feedback.

EVALUATION

We evaluated the system to find out if and how IntentStreams supports parallel browsing and branching behavior. IntentStreams was compared against a baseline system with an interface similar to a traditional Google search interface. Our hypothesis was that, compared to the baseline, IntentStreams generates (1.) more parallel streams, (2.) more revisits, and (3.) more branches. We used the following metrics: number of parallel streams, number of revisits, and number of branches. In the baseline, the number of parallel streams denotes the number of tabs opened, a revisit indicates returning to an already open tab, and a branch denotes a query updated after a revisit. In IntentStreams, a revisit occurs when a user performs certain activities (opening an article, weight change) on a previously created stream. A branch occurs when a new stream is created from an existing one. That includes both creating a new query by dragging a keyword or updating the existing stream by modifying the weights of its keywords.

Method

We evaluated the system with 13 volunteers (4 female). The participants' age ranged from 19 to 36 with mean of 28.4 ($SD = 4.05$). Their levels of education were: 8% PhD, 46% Master, 38% Bachelor, 8% High School. Each participant received two movie tickets for their participation. We used a within-subject design, where participants were asked to perform two tasks, one with IntentStreams and one with the baseline. We counterbalanced by changing the order in which the two tasks were performed and the order in which the two systems were used.

The task was set in an essay writing scenario and formulated as follows: *You have to write an essay on recent developments of X where you have to cover as many subtopics as possible. You have 20 minutes to collect the material that will provide inspiration for your essay. You have additional 5 minutes to write your essay.* The two tasks performed by the participants covered two topics: (1.) NASA, and (2.) China Mobile.

Experiments were run in a laboratory on a laptop with OS X operating system. Each participant signed a consent form. To determine the eligibility, we asked candidates how familiar they were with each chosen topic on a 1-5 scale, where 1 means "no knowledge" and 5 means "expert knowledge". Only those with a score lower than 3 were considered eligible. Before the experiment, participants received detailed instructions and performed a 5-minute training session.

To evaluate the system, we connected it to a news repository of English language editorial news articles crawled from publicly available news sources from September 2013 to March 2014. The database contains more than 25 million documents. The documents were originally collected for monitoring media presence of numerous interested parties, and hence the collection has wide topical coverage. All the documents were preprocessed by the Boilerpipe tool [19] and the keyphrases were extracted with the Maui toolkit [25].

The baseline system was connected to the same news repository. In the baseline system, users could type queries and receive a list of relevant news articles. To start a new parallel query, a new tab had to be opened.

FINDINGS

Table 1 shows the results of the log analysis. In the 20–minute long sessions, IntentStreams on average generated 7.84 more queries ($SD = 7.27$), 6.38 more parallel streams ($SD = 4.03$), 4.54 more revisits ($SD = 4.52$), and 3.62 more branches ($SD = 4.01$). A paired t-test indicates that all those differences are statistically significant ($p < 0.01$).

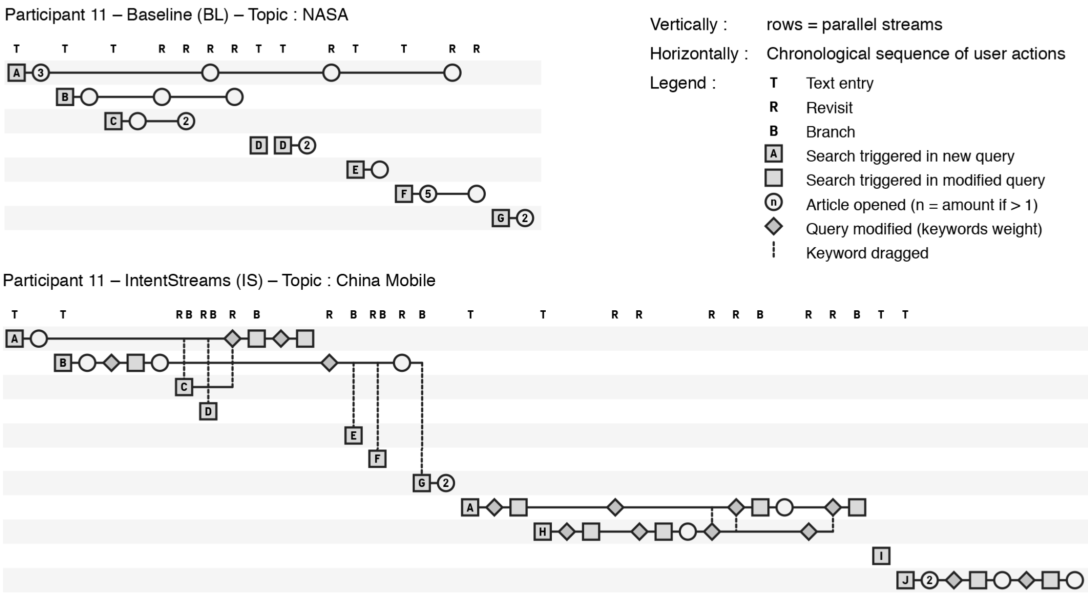


Figure 2. Example of branching behavior from the case study: top - Baseline; bottom - IntentStreams.

Parallel search supported in IntentStreams. Results show that users created more parallel streams than opened new tabs. While the system allows the creation of parallel streams, the users revisit earlier ones consistently, which denotes parallel search behavior. In fact, revisits are higher in the IS condition.

Branching supported in IntentStreams. In IntentStreams, more queries and parallel streams were created through branching. Figure 2 presents a visual representation of a participant's search behavior, showing the difference between the linear search behavior in the baseline and the more articulated search behavior in IntentStreams.

Further, IntentStreams supports more exploration. In IS, more exploration of the information space was done as can be seen from the higher number of queries.

CONCLUSIONS

We introduced the IntentStreams system for exploratory search of news based on parallel visualization of smart search streams. It models each search stream by an intent model, allows rapid tuning by feedback to keywords, and allows rapid initiation of new streams by keyword interaction without typing. Initial experiments show that users take advantage of the rich parallel search opportunities and engage in much stronger parallel browsing and branching behavior than in a traditional system.

This is an important finding as current browsing and searching behavior is already characterized by multitask search (in the same query field users alternate tasks [22, 21]), parallel browsing (users browse on parallel tabs or windows [14]), and engage in branching (a new tab or window is created from a link or result of a previous window or tab [13]). Branching has been shown to be more important in informational browsing than navigational search [14]. The approach proposed in IntentStreams can be incorporated into other search interfaces to provide an effective way to branch search.

Table 1. Comparison between IntentStreams (IS) and the baseline (BL). The number of queries, parallel streams, revisits, and branches, for each participant P1,...,P13.

	queries		par. streams		revisits		branches	
	BL	IS	BL	IS	BL	IS	BL	IS
P1	5	5	2	5	0	7	0	1
P2	5	4	1	2	1	0	1	0
P3	7	17	1	12	1	6	1	4
P4	9	11	1	6	0	2	0	2
P5	14	18	1	12	0	9	0	7
P6	1	8	1	5	0	0	0	0
P7	12	18	7	14	5	7	2	3
P8	22	26	3	12	6	15	1	8
P9	6	18	6	12	4	4	0	0
P10	8	11	7	11	6	5	0	0
P11	8	21	7	11	7	12	0	8
P12	16	35	11	14	3	14	1	9
P13	3	26	3	18	0	11	0	11

ACKNOWLEDGMENTS

This work has been partially supported by TEKES (Re:Know and DIGILE programme D2I) and the Academy of Finland (decisions 251170, 252845, 266969, 278090, and 283311).

REFERENCES

1. Ahlberg, C., and Shneiderman, B. Visual information seeking: Tight coupling of dynamic query filters with

starfield displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (New York, NY, USA, 1994), 313–317.

2. Ahn, J., and Brusilovsky, P. Adaptive visualization for exploratory information retrieval. *Information Processing & Management 49*, 5 (2013), 1139–1164.

3. Ahn, J.-w., and Brusilovsky, P. Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization 8*, 3 (2009), 167–179.

4. Auer, P. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research 3* (2002), 397–422.

5. Bates, M. J. What is browsing–really? a model drawing from behavioural science research. *Inf. Res. 12*, 4 (2007).

6. Bilenko, M., and White, R. W. Mining the search trails of surfing crowds: identifying relevant websites from user activity. In *Proceedings of the 17th international conference on World Wide Web*, ACM (2008), 51–60.

7. Chau, D. H., Kittur, A., Hong, J. I., and Faloutsos, C. Apolo: Making sense of large network data by combining rich user interaction and machine learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, ACM (New York, NY, USA, 2011), 167–176.

8. Chirita, P.-A., Firan, C. S., and Nejdl, W. Personalized query expansion for the web. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM (2007), 7–14.

9. Głowacka, D., Ruotsalo, T., Konuyshkova, K., Kaski, S., Jacucci, G., et al. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, ACM (2013), 117–128.

10. Głowacka, D., and Shawe-Taylor, J. Content-based image retrieval with multinomial relevance feedback. In *Proc. of ACML* (2010), 111–125.

11. Havre, S., Hetzler, E., Perrine, K., Jurrus, E., and Miller, N. Interactive visualization of multiple query results. In *Proceedings of the IEEE Symposium on Information Visualization 2001*, IEEE Computer Society (Washington, DC, USA, 2001), 105.

12. Hearst, M. A. *Search User Interfaces*, 1st ed. Cambridge University Press, New York, NY, USA, 2009.

13. Huang, J., Lin, T., and White, R. W. No search result left behind: Branching behavior with browser tabs. In *Proceedings of the fifth ACM international conference on Web search and data mining*, ACM (2012), 203–212.

14. Huang, J., and White, R. W. Parallel browsing behavior on the web. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, ACM (2010), 13–18.

15. Käki, M. Findex: Search result categories help users when document ranking fails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (New York, NY, USA, 2005), 131–140.

16. Kelly, D., and Fu, X. Elicitation of term relevance feedback: An investigation of term source and context. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM (New York, NY, USA, 2006), 453–460.

17. Kelly, D., Gyllstrom, K., and Bailey, E. W. A comparison of query and term suggestion features for interactive searching. In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM (New York, NY, USA, 2009), 371–378.

18. Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., and Jacucci, G. Designing for exploratory search on touch devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI 2015, ACM (2015).

19. Kohlschütter, C., Frankhauser, P., and Nejdl, W. Boilerplate detection using shallow text features. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ACM (2010).

20. Kules, W., Wilson, M. L., Schraefel, M. C., and Shneiderman, B. From keyword search to exploration: How result visualization aids discovery on the web. Tech. rep., University of Southampton, 2008.

21. Lucchese, C., Orlando, S., Perego, R., Silvestri, F., and Tolomei, G. Identifying task-based sessions in search engine query logs. In *Proceedings of the fourth ACM international conference on Web search and data mining*, ACM (2011), 277–286.

22. Lucchese, C., Orlando, S., Perego, R., Silvestri, F., and Tolomei, G. Discovering tasks from search engine query logs. *ACM Transactions on Information Systems (TOIS) 31*, 3 (2013), 14.

23. Marchionini, G. Exploratory search: From finding to understanding. *Communications of the ACM 49*, 4 (Apr. 2006), 41–46.

24. Matejka, J., Grossman, T., and Fitzmaurice, G. Citeology: Visualizing paper genealogy. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, ACM (New York, NY, USA, 2012), 181–190.

25. Medelyan, O., Frank, E., and Witten, I. H. Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (2009), 1318–1327.

26. Micarelli, A., Gasparetti, F., Sciarrone, F., and Gauch, S. Personalized search on the world wide web. In *The Adaptive Eeb*. Springer, Berlin, Heidelberg, 2007, 195–230.

27. Pirolli, P., and Card, S. Information foraging in information access environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '95, ACM Press/Addison-Wesley Publishing Co. (New York, NY, USA, 1995), 51–58.

28. Ruotsalo, T., Jacucci, G., Myllymäki, P., and Kaski, S. Interactive intent modeling: Information discovery beyond search. *Communications of the ACM 58*, 1 (Dec. 2014), 86–92.

29. Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K., Kosunen, I., Reijonen, A., Myllymäki, P., Jacucci, G., et al. Directing exploratory search with interactive intent modeling. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, ACM (2013), 1759–1764.

30. Stasko, J., Görg, C., and Liu, Z. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization 7*, 2 (2008), 118–132.

31. Teevan, J., Dumais, S. T., and Liebling, D. J. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, ACM (2008), 163–170.

32. Terveen, L., Hill, W., and Amento, B. Constructing, organizing, and visualizing collections of topically related web resources. *ACM Transactions on Computer-Human Interaction 6*, 1 (1999), 67–94.

33. Venna, J., Peltonen, J., Nybo, K., Aidos, H., and Kaski, S. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research 11* (2010), 451–490.

34. Verbert, K., Parra, D., Brusilovsky, P., and Duval, E. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, ACM (2013), 351–362.

Publication III

Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, (CHIIR '17). ACM, New York, NY, USA, pp. 57-66, 2017. DOI: 10.1145/3020165.3020174, March 2017.

Visual Re-Ranking for Multi-Aspect Information Retrieval

Khalil Klouche[1,3], Tuukka Ruotsalo[2], Luana Micallef[2]
Salvatore Andolina[2], Giulio Jacucci[1,2]
[1]Helsinki Institute for Information Technology HIIT, Department of Computer Science,
University of Helsinki, PO Box 68, 00014 University of Helsinki, Finland
[2]Helsinki Institute for Information Technology HIIT, Aalto University,
PO Box 15600, 00076 Aalto, Finland
[3]Aalto University, School of Arts, Design and Architecture, Media Lab Helsinki,
Hämeentie 135 c, 00560 Helsinki, Finland
[1]first.last@helsinki.fi, [2]first.last@aalto.fi

ABSTRACT

We present visual re-ranking, an interactive visualization technique for multi-aspect information retrieval. In multi-aspect search, the information need of the user consists of more than one aspect or query simultaneously. While visualization and interactive search user interface techniques for improving user interpretation of search results have been proposed, the current research lacks understanding on how useful these are for the user: whether they lead to quantifiable benefits in perceiving the result space and allow faster, and more precise retrieval. Our technique visualizes relevance and document density on a two-dimensional map with respect to the query phrases. Pointing to a location on the map specifies a weight distribution of the relevance to each of the query phrases, according to which search results are re-ranked. User experiments compared our technique to a uni-dimensional search interface with typed query and ranked result list, in perception and retrieval tasks. Visual re-ranking yielded improved accuracy in perception, higher precision in retrieval and overall faster task execution. Our findings demonstrate the utility of visual re-ranking, and can help designing search user interfaces that support multi-aspect search.

Keywords

Information visualization; information retrieval; multi-aspect search; multi-dimensional ranking

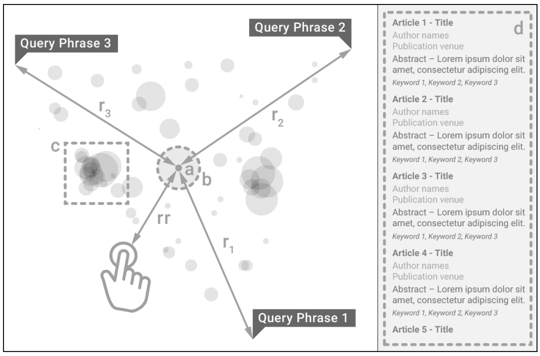


Figure 1: Interactive relevance map visualization. (a) Position of a document marker is computed as a weighted linear combination of relevance to individual query phrases $r1, r2, r3$. (b) Radius of a document marker encodes the overall relevance of the corresponding document to all query phrases. (c) Opacity encodes the density of document mass in a certain position of the 2D plane. (d) The result list can be re-ranked by relevance and the distance to the selected position rr.

1. INTRODUCTION

Multi-aspect search refers to activities in which the information need of the user consists of more than one aspect or query simultaneously. Such situation arises in contexts such as exploratory search, item selection and multi-criteria decision making. In exploratory search activities, the user's goal is not clearly defined, and the information space is usually unfamiliar to the user. In such scenarios, the user might start from a small set of notions, with the intent of learning and making sense of the related document space. In this case, conventional result lists offer little insight of the data and nothing indicates how the given results relate to the multiple aspects of the query. For example, a user looking for recent literature on physiological measurements might want to search for aspects such as 'Electroencephalography', 'Electrodermal Activity', 'Electromyography' and quickly be able to assess how the result space is distributed and how the retrieved documents relate to each aspect.

Item or product selection is currently widely supported by faceted search and search result clustering. Such systems are widespread in e-commerce and library catalogs. These techniques allow the user to investigate the results through the use of multiple filters, but they offer limited support for perceiving the result space and weighting the aspects accordingly. Conventional query-based search tools usually visualize results as a one-dimensional ranked list, and offer limited support for multi-aspect retrieval. Another example is multi-criteria decision making, a well researched process that often requires multi-aspect search [8]. Take the example of a user looking online for a new car. Usual faceted tools allow her to select filters to narrow down the offering: e.g., a manufacturer, a price range, a fuel type. Such criteria require the user to have

CHIIR '17, March 07-11, 2017, Oslo, Norway

DOI: http://dx.doi.org/10.1145/3020165.3020174

a specific goal in mind, whereas a typical user would be inclined to come up with more vague criteria such as: *good gas-mileage* (no specific threshold in mind), *family-friendly*, and/or *fun to drive*. Such criteria are not binary, and the user can expect to find on the market several satisfying solution with different tradeoffs, instead of one ideal car. On the other hand, looking for such criteria using one single unified query on a conventional search engine returns a list of results that does not reflect the user's preferences and does not allow for conscious tradeoffs.

In all these cases, the user should be able to quickly assess distribution of the results with respect to how they relate to each researched aspect, and then be able to rapidly inspect them, which is possible if the user 1) perceives the distribution to understand which parts of the result space contain interesting information (i.e. what are the tradeoffs between the query phrases) and 2) is able to determine the tradeoff rapidly using the visualization.

We present a visual re-ranking technique that uses multi-dimensional ranking and two-dimensional interactive visualization. Inspired by earlier work on visual information retrieval and seeking [33, 2], the technique allows the user to perceive the relevance distribution with respect to multiple query phrases by using a *relevance map* visualization. A novel feature of this technique is that it allows the user to investigate specific areas on the map by re-ranking the results through pointing at the map. The method estimates document relevance with respect to user-specified query phrases in a multi-dimensional space in which the query phrases define the dimensionality. The method then computes a layout for the documents on a two-dimensional plane where relative distances of document markers to each query phrases are defined by their respective relevance, overall relevance of each document is visualized as the radius, and higher document density translates in darker areas. (see Figure 1). The visualization allows the user to perceive how the result space is populated with respect to both density and relevance to some query phrases.

Rather than relying only on a one-dimensional ranking algorithm to select the documents most relevant to a query, the role of the system is to organize and present information about many documents and multi-dimensional query phrases in a way that makes comparison possible. Re-ranking by pointing allows users to rank documents with respect to relative relevance weights to the query phrases. For example, expressing that a user wants the ranking to be based a little on both query phrases *interaction* and *interfaces*, but mainly on the phrase *design* can be done simply by pointing to an area on the map that is inside a triangle of the query phrases but closer to the concept *design*.

The approach was evaluated in a controlled laboratory study with 20 participants performing two tasks: perception and retrieval. In the perception task, participants were asked to find out how a document space was populated and organized with respect to specific topics, such as whether there was more research about *interaction* or *design*. In the retrieval task, the participants were asked to find documents with varying relevance to several topics, such as a document that was mainly related to design, but slightly related to interaction and interfaces.

Our results show significant improvement in task completion time as well as improved accuracy in perception, and improvement in task completion time in retrieval, without compromising effectiveness measured as the quality of the task outcome. These results suggest that relevance mapping and re-ranking is effective in cases when the initial one-dimensional result list is not enough for the user to analyze the information.

The contributions of this paper are: (1) We present a visual re-ranking approach to multi-aspect information retrieval in which users can perceive the result space and rapidly re-rank the result list by pointing to the visualization. (2) We demonstrate that users can complete perception and re-ranking tasks significantly faster without compromising the effectiveness. (3) While different approaches for search result visualization have been proposed in the past, up to our knowledge, this is the first study that empirically verifies the benefits of interactive visualization for multi-aspect information retrieval.

2. RELATED WORK

2.1 Visual Information Retrieval and Seeking

Information spaces can be huge and thus hard to comprehend. However, visualizing the space and allowing the user to directly interact with and manipulate objects in the space facilitates comprehension. For instance, when the results of actions are shown immediately and when typing is replaced with pointing or selecting, exploration and retention increase while errors decrease [46]. For information seeking, the following visualization and interaction features are of particular importance [43]: (a) dynamic querying for rapid browsing and filtering to view how results change; (b) a starfield display for the immediate, continuous, scalable display of result sets as different queries are processed; (c) tight coupling of queries to easily use the output of one query as input to another [1]. For instance, a user study indicates that dynamic querying significantly improves user response time and enthusiasm. Using such techniques, systems like FilmFinder [1] support querying over multiple varying attributes such as time, while showing the changing query results in the context of the overall data. User studies also indicate that user interfaces that show the result list together with an overview of the result categories encourage a deeper and more extensive exploration of the information space [25], especially when the system allows relevance feedback to be given on such categories to direct the exploration [40, 39].

2.2 Document Collection Visualization

Various visualizations have been proposed for large document collections [24]. Most of these techniques adopt the visual information seeking mantra [44] to provide an overview at first and details only on demand. The documents are often visualized on a 2D plane, in the form of a map based on a similarity metric. Higher-level entities, such as topics, are also displayed on the map for immediate and better understanding of the document space organization.

Document Atlas [12] uses Latent Semantic Indexing and multi-dimensional scaling (MDS) to extract semantic concepts from the text and position the documents with respect to the concepts. Document densities around concepts are visualized as a heat map. On mouse hover, common keywords in the area are listed, and on zoom in, more details are shown.

Self-Organizing Maps have also been used by systems like WEBSOM [20] and Lin's maps [26] to position the documents on the 2D plane. WEBSOM also suggests areas in the map that could be relevant to the user's search query. Lin's maps are further split up into regions whose area indicates the number of documents with specific related terms.

Other techniques visualize the documents as glyphs to indicate additional inter-document relationships and metadata on the map (e.g., [38, 29]). Various metaphors have also been adopted; examples include the terrain metaphor, in which dense regions in the map are seen as mountains with valleys in between [9, 49]; the galaxy metaphor, in which documents are seen as stars in different constellations (document clusters) [16]; and the physical metaphor, in which documents are considered to be moving particles and the inter-particle forces move similar documents closer to each other and dissimilar documents apart [10]. Visualizations with two dimensions and meaningful axes (e.g., categories vs. hierarchies [45], query results vs. query index [6, 23], production vs. popularity [1]) have also been proposed.

These visualizations provide an overview of the entire document collection, but they do not allow the user to direct and focus the exploration as required. A user-driven rather than a data-driven technique could be more helpful when searching for documents relevant to multiple keywords. To that end, such a technique should visualize the ranking of documents with respect to multiple keywords so the user can easily judge the relevance of documents to each of the keywords of interest [33]. However, most of the current techniques only visualize whether a document is relevant or not to a keyword using set visualizations [4], without showing the document's degree of relevance to each keyword.

2.3 Multi-Aspect Search

In multi-aspect search the information need of the user consists of more than one aspect or query simultaneously. As a consequence, an item in a collection needs to be ranked differently based on its multiple attributes. The Graphics, Ranking, and Interaction for Discovery (GRID) principles and the corresponding rank-by-feature framework state that interactive exploration of multi-dimensional data can be facilitated by first analyzing one- and two-dimensional distributions and then by exploring relationships between the dimensions, using multi-dimensional rankings to set hypotheses and statistics to confirm them [41]. However, comparing, analyzing and relating different ranks is difficult and requires an interactive visualization that supports the various requirements identified by Gratz et al. [13].

Multi aspect search support is provided in Song et al. [47], with the proposal of a strategy for multi-aspect oriented query summarization task. The approach is based on a composite query strategy, where a set of component queries are used as data sources for the original query. Similarly Kang et al. [19] propose a multi-aspect relevance formulation, but in the context of vertical search.

LineUp [13] is an interactive visualization that uses bar charts to support the ranking of objects with respect to multiple heterogeneous attributes. Stepping Stones [11] visualizes search results for a pair of queries, using a graph to show relationships between the two sets of results. Sparkler [14] allows to visually compare results sets for different queries on the same topic.Tilebars [15] visualizes the frequency of different words in various sections of documents as a heat map and ranks the documents accordingly. Similarly, HotMap uses a two-dimensional grid layout to augment a conventional list of search results with colors indicating how hot (relevant) specific search terms are with respect to the document [18]. Ranking cube [50] is a novel rank-aware cube structure that is capable of simultaneously handling ranked queries and multi-dimensional selections. RankExplorer [42] uses stack graphs for time-series data. Techniques for incomplete and partial data have also been proposed [22]. TreeJuxtaposer [31] was primarily devised to compare rankings.

For document collections, the vector space model could be used, such that each document and search query is a vector in a multi-dimensional space, each axis is a term, and the document position is determined by the frequencies of each term in that document (e.g., [36]). Visualizations of such a model could aid understanding of the document space, but more research is required, particularly for user-driven approaches that allow the user to specify the dimensions of interest [33].

2.4 User-driven Visualization

VIBE [33] is one of the most well-known user-driven multi-dimensional ranking visualization for large document collections. To indicate the subspace of interest, the user first enters two or more query terms, known as "points of interest" (POIs). POIs are then shown (as circles) on a 2D plane, together with documents (as rectangles) related to at least one POI, forming a map. The position of each rectangle indicates the relevance of the corresponding document to each of the POIs. The size of a rectangle indicates the relevance of that document to the search query. Citation details of documents selected from the map are listed; clicking on an item in the list opens the full document. Any time a POI is added, removed or moved, the map is updated accordingly. However, regions of the map with numerous close-by documents are not easily detectable because the rectangles are not color filled; using semi-transparent color filled shapes reduces overplotting [28] and facilitates perceptual ordering of different regions in the map by their density [27]. Also, documents are not re-ranked as the user navigates over the map.

Variants of VIBE include: WebVIBE [30], in which POIs act like magnets that attract documents containing related terms; VR-VIBE [7], which visualizes the space in 3D (for more space to view documents between POIs) and depicts relevance by color; and Adaptive VIBE [3], in which POIs are query terms (as in VIBE) but also user profile terms that are automatically extracted from user notes.

Similar to VIBE, GUIDO [32], DARE [53] and TOFIR [52] also allow users to specify POIs and display documents based on their relevance to the POIs. However, in GUIDO each POI is an axis (not an icon on a 2D plane) and documents are positioned based on their absolute rather than relative distances from the POIs. In DARE and TOFIR, relevance to POIs is indicated by both distance and angle.

Other user-driven systems, like combinFormation [21], TopicShop [5] and InfoCrystal [48], retrieve and display search results related to user-defined keywords but do not visualize the results' multi-dimensional ranks. Similarly, HotMap [18] supports a weighted re-ranking of the search results, but without leveraging a graphical interactive approach for specifying the weights. WordBars [17] also supports re-ranking of the search results, but uses additional terms extracted from the search results rather than relying on the query terms.

While similar techniques of mapping data to 2D visualization for better user interpretation have been proposed, the current research lacks understanding on (1) how useful these are for the user and (2) whether they lead to quantifiable benefits in specific tasks related to search activity. This work is the first to demonstrate a technique where the visualization can be effectively used for re-ranking search results. It is also the first that empirically verifies that users perceive the document space faster and are able to execute retrieval faster without compromising the quality of retrieved information.

3. RELEVANCE MAPPING

The method for relevance mapping is first illustrated with an overview from the user perspective. Then the computation of the layout and document visualization is explained.

3.1 Overview

Figure 2 shows an example of the relevance map visualization. Here, a user investigates a document space delimited by three query phrases with corresponding markers on the map: *design*, *interaction* and *interface*. A fourth query marker, *exploration*, is greyed out because it has been disabled to permit a temporary focus on the three remaining query markers. The user has positioned the

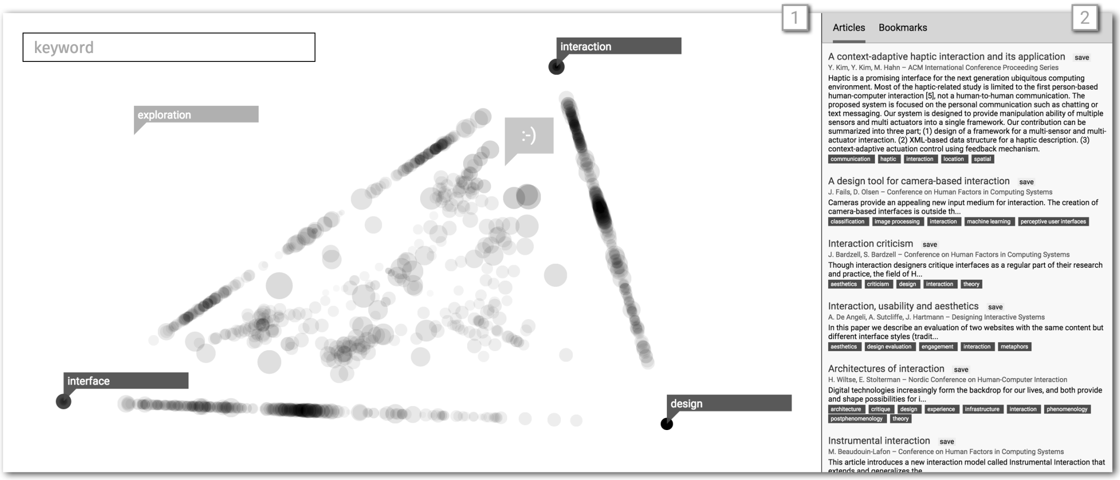


Figure 2: The relevance map (1) displays documents in relation to multiple query phrases, displayed as red text labels. Here, a fourth query phrase is greyed out (disabled). The exploration cursor (in blue) is located at the user-specified position to be used for the re-ranking. Red document markers indicate the position of articles currently on display in the result list (2).

pointer (blue flag with a smiley face) close to *interaction* to investigate a collection of documents highly related to *interaction* and more loosely related to *interface* and *design*. As a result, the list shows articles ranked with a specific focus on the selected area.

Query markers are created by inputting keywords in the query box in the top left. Each query marker can be activated or disabled by clicking it. Documents returned by the system are visualized on the map as semi-opaque dots scattered between the query markers with respect to their individual relevance. The overall relevance of a document is indicated by the radius of the dot. The partial opacity translates overlapping into a darkened tint that cues the user on the number of document markers in any given area. Query markers can be moved/dragged around on the map, which updates the position of the document markers. The position of the pointer can be positioned by dragging or tapping on the map. Any change in the pointer position or query marker organization triggers a re-ranking of documents based on their overall relevance and proximity to the pointer.

The ranked articles appear in a conventional one-dimensional list layout in the result list (2). Documents being displayed in the result list are shown as red dots on the map. The result list is scrollable. Each document is displayed with its title, authors, publication venue, abstract and keywords. Abstracts are first shown partially but can be displayed in full at a click or a tap. Keywords are interactive, as they can be added to the map as new query markers on a tap.

3.2 Layout

The data used to compute the relevance map layout consists of a set of m query phrases $q_{1...m} \in Q$, a set of k documents $d_{1...k} \in D$ and relevance estimates $r_{1...k} \in R$ for each of the k documents according to each of the m query phrases.

Each query marker and each document marker has a position on the plane, pos_{q_x}, pos_{q_y} and pos_{d_x}, pos_{d_y} respectively. The position of each query phrase marker is defined by the user by moving it to the desired position on the plane. The position of each of the document markers is computed as a weighted linear combination of the relevance scores to each query phrase and the relative position of the query marker. Intuitively, document markers are positioned proportional to their relevance to each of the query phrases. Formally, the position of an jth document marker on dimension dim is:

$$pos_{d_{j_{dim}}} = \frac{\sum_{i}^{|Q|} r_{q_i d_j} \cdot pos_{q_{dim}}}{|Q|} \quad (1)$$

so that $pos_{d_{i_{dim}}}$ is the coordinate of document d_i with respect to dimension dim. On a two-dimensional plane dim can be x or y. The relevance estimation $r_{q_i d_j}$ of a document to a query phrase is explained in the next section.

3.3 Document Marker Visualization

The radius of the document marker is directly the relevance $r_{q_i d_j}$. That is, the size of the dot is defined by the relevance.

The opacity of overlapping document markers is used to visualize the density of the document mass in a particular position on the plane. We use a standard computation of opacity [35] in which opacity of o of a pixel on the plane is computed as:

$$o = 1 - (1 - f)^n \quad (2)$$

where n is the number of overlapping layers and f is a constant setting of an opacity effect of an individual layer and was set to $f = 0.95$.

4. RELEVANCE ESTIMATION

The relevance estimation used in ranking and computing the document marker layout and size are explained in this section.

4.1 Relevance Estimation

Given the document collection and a set of query phrases that specify the multiple dimensions to be used in ranking and visualization, the relevance estimation method results in a set of probabilities $r_{1...k} \in R$ for each document d of k documents in the collection according to each query phrase $q_{1...m} \in Q$.

To estimate the probabilities from the query phrases Q and documents D, we utilize the language modeling approach of information retrieval [34]. We use a multinomial unigram language model. The vector Q of query phrases is treated as a sample of a desired document, and document d_j is ranked according to a query phrase q_i by the probability that q_i would be generated by the respective language model M_{d_j} for the document; with the maximum likelihood estimation we get

$$P(q|M_{d_j}) = \prod_{i=1}^{m} \hat{P}_{mle}(q_i|M_{d_j})^{w_i}, \tag{3}$$

where w_i is the weight of each of the query phrases and is set as $w_i = \frac{1}{|Q|}$ as default. In case of interactive re-ranking w_i is weighted based on user interactions as explained in the next section.

To estimate the relevance $r_{q_i d_j}$ of an individual document d_j with respect to an individual dimension defined by each query phrase q_i and avoid zero probabilities, we then compute a smoothed relevance estimate by using Bayesian Dirichlet smoothing for the language model so that

$$r_{q_i d_j} = P_{mle}(q_i|M_{d_j}) = \frac{c(q_i|d_j) + \mu p(q_i|C)}{\sum_k c(q|d_j) + \mu}, \tag{4}$$

where $c(d_i|d_j)$ is the count of a query phrase q_i in document d_j, $p(q_i|C)$ is the occurrence probability (proportion) of a query phrase q_i in the whole document collection, and the parameter μ is set to 2000 as suggested in the literature [51].

4.2 Ranking

Given the probability estimates for each of the documents, we apply a probability ranking principle [37] to rank the documents in descending order of their probabilities for the query phrases. These are then used to compute the total ordering of the document list. The top-k ranking computation remains efficient by making use of priority queue with complexity log(k) of k search results with presorted inverted index.

The user can interactively re-rank the result list by selecting a point on the relevance map. The point for the desired re-ranking is defined by its two-dimensional coordinates rr_x and rr_y with respect to the two-dimensional coordinates of the query markers $pos_{q_{ix}}$ and $pos_{q_{iy}}$ for the $i = 1 \ldots |Q|$ query phrases.

The re-rank weighting for an ith query marker is computed as the Euclidean distance between the $pos_{q_{ix}}$ and $pos_{q_{iy}}$ and the rr_x and rr_y. Formally,

$$w_i = \frac{\sqrt{(pos_{q_{ix}} - rr_x)^2 + (pos_{q_{ix}} - rr_y)^2}}{\sum_i^{|Q|} q_i}, \tag{5}$$

The re-ranking of the documents is then computed using these distances by Formula 3 by setting the weight w_i accordingly. Intuitively, the distance from the query marker is used as the importance of the query phrase in the ranking of the documents.

5. EXPERIMENTS

The current research lacks understanding on the end-user benefits of interactive visualization in multi-aspect search scenarios. The perceived simplicity and overall familiarity of well-studied conventional search system interfaces – like the current de-facto search interface with typed query and a ranked result list – have not been challenged in experiments that measure the quantifiable benefits of task completion time and effectiveness.

We conducted a controlled laboratory experiment in which the relevance mapping and re-ranking were compared to a conventional ranked list visualization in two basic tasks that searchers have to perform when using an information retrieval system: perception and retrieval.

The perception task sought understanding on the benefits of the visualization in perceiving the distribution and density of resulting documents with respect to the multi-aspect query phrases. The retrieval task sought understanding on the benefits of the visualization in re-ranking the results according to a user specified distribution over the importance of the different query phrases (see Figures $3b$,$3c_1$, and $3c_2$). The benefits were measured with respect to task completion time and effectiveness (quality of the perception or retrieval). The following subsections explain the details of the experiments.

5.1 Hypotheses

The study tested the following four hypotheses:

- $H1$: Efficient perception hypothesis: The relevance map allows faster perception of the result set.
- $H2$: Efficient retrieval hypothesis: The relevance map allows faster retrieval of relevant information.
- $H3$: Effective perception hypothesis: The relevance map allows more accurate perception of the result set.
- $H4$: Effective retrieval hypothesis: The relevance map allows retrieval of more highly relevant information.

5.2 Experimental Design

The experiment used a 2×2 within-subjects design with two search tasks and two systems. The conditions were counterbalanced by varying the order of the systems and tasks.

5.3 Baseline

A baseline system, shown in Figure $3a$, was implemented to enable comparability and as to ensure that the evaluation revealed the effects solely on the features enabling relevance mapping and re-ranking. The baseline used the same data collection as well as the same document ranking model. All retrieved information in the baseline system was displayed with a ranked list layout. The baseline did not feature a relevance map, and the ranking was based on a single query at a time. The baseline was using the same hardware, i.e. a multi-touch-enabled desktop computer with a physical keyboard.

5.4 Tasks

The experiment consisted of two tasks, perception and retrieval, which are explained below and exemplified in Figures $3b$,$3c_1$, and $3c_2$. Both tasks used a common set of four topics, either (1) interaction, tabletop, tangible, and prototyping, or (2) surfaces, exploration, visualization, and sound. The two set of topics were formed by two researchers who were experts on human-computer interaction. The same researchers were then asked to assess the task outcomes of the participants.

5.4.1 Perception Task

The perception task aimed to measure task completion time and effectiveness, to help understand how a document space is populated and organized with respect to specific query topics. Participants were asked the two following questions: (1) "Out of the 4 topics provided, which 2 topics are related to the highest amount of

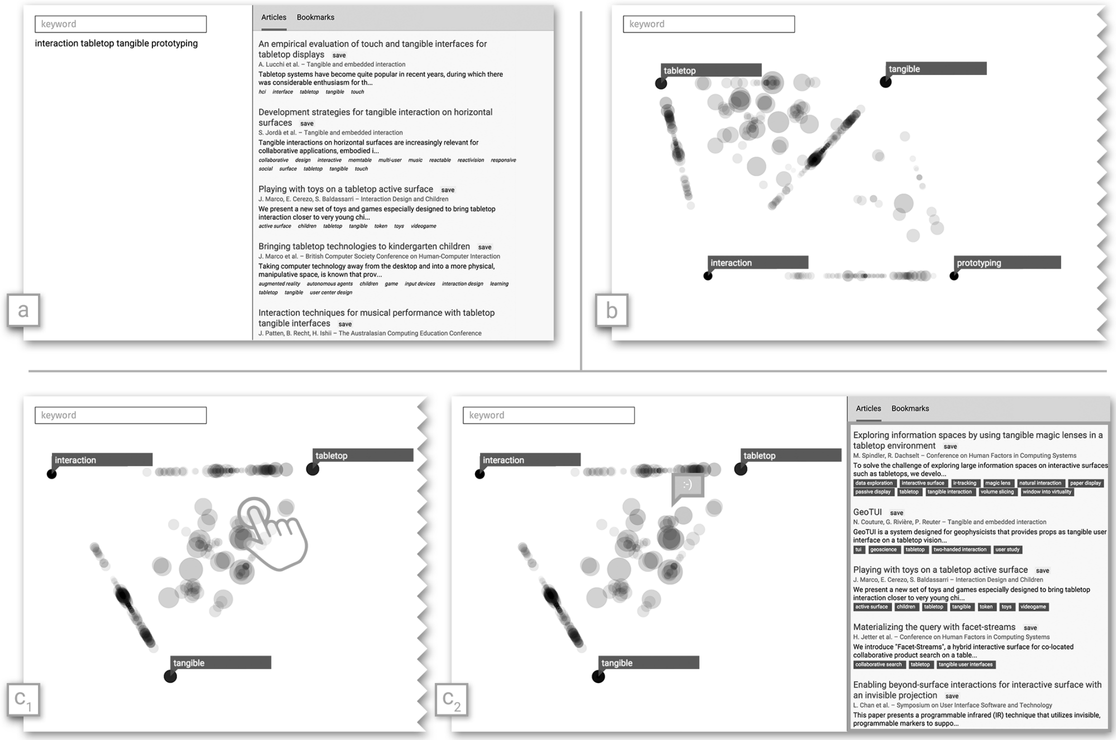


Figure 3: In the perception task, participants must identify the two and three keywords out of four that are the most related to relevant information. In the baseline (a), they must skim through the ranked list of results to infer the most prevalent keywords from the top articles. Using the relevance map, they must interpret the distribution of document markers. In the retrieval task, participants must find an article that shows a high relevance to one keyword (say, *tabletop*), and a lesser relevance with two other keywords (say, *tangible* and *interaction*). Using the baseline (a), they must query the three keywords, then find a fitting article in the result list. Using the relevance map, they point (by tapping on the touch-enabled monitor) at an area between the three keywords (c1), somewhere closer to *tabletop* than *tangible* or *interaction*, which triggers a re-ranking of retrieved articles based on the selected position (c2). The participant should be able to select one of the top articles as a fitting task outcome.

relevant documents?", and (2) "Out of the 4 topics provided, which 3 topics are related to the highest number of relevant documents?".

An example visualization from which the user had to select the topics is shown in Figure 3*b*. In that case, we can see that the space delimited by *tabletop*, *tangible*, and *interaction* is the most densely populated through sheer amount of document markers, making them part of the answer. To find the two keywords, they must then compare pair-wise document density by focusing on the edges between query phrase markers, with a slight but noticeable lead in density (encoded as darkness) between *tangible* and *interaction*.

5.4.2 *Retrieval Task*

The retrieval task aimed to measure task completion time and effectiveness in finding documents with varying multi-dimensional relevance toward several topics. Participants were given the following instruction: "Find one article that is highly relevant to ‘Topic A’ and slightly related to ‘Topic B’ and ‘Topic C’.". The task was then repeated one more time with a different topic priority: "Find one paper that is highly relevant to ‘Topic B’ and slightly related to ‘Topic A’ and ‘Topic C’."

An example sequence of a visualization, user pointing to the visualization to re-rank the document list from which the user had to select the documents is shown in Figures $3c_1$ and $3c_2$.

5.5 Measures

We used two performance measures: task completion time and effectiveness. *Task Completion Time* measured the time required to complete the task. *Effectiveness* measured the quality of the task outcome.

5.5.1 *Task Completion Time*

Task completion time was computed directly as the duration in seconds from the beginning of the task to the completion of the task.

5.5.2 *Effectiveness*

Effectiveness was computed differently for the two tasks and the corresponding ground truths for the task outcomes were defined differently.

In the perception task, effectiveness was measured as the accuracy of the participants answer. The ground truth was available from the relevance estimation and was computed as a sum of the relevance scores associated to each query phrase representing the topic. The topics were then ordered based on the sum of relevance scores and the top 2 and top 3 topics corresponding to the task description were selected as the ground truth to which each answer was then compared. Accuracy was computed for each answer, resulting in a grade of 1 for a match, 0 for a mismatch, and – in the case two topics selected out of four – 0.5 for a partial match. Each participant having returned two answers, effectiveness was then measured as the mean of both grades.

In the retrieval task, effectiveness was measured as precision on the documents selected by the participants. All documents chosen by any of the participants in any of the two system conditions were pooled. Two experts then assessed the actual relevance of each document to each topic. The experts being authors of the experiment design and having themselves devised the topics, potential bias in the assessment was addressed by following a strict double-blind procedure (i.e. experts had no knowledge of the participant, the system or concurrent assessment) and balancing the use of each set of topics across both conditions. The experts assigned for each document a grade between 0 (non-relevant) and 5 (highly relevant) to each of the topics, which were then averaged (mean) into a final grade. The topic defined as highly relevant was given a double coefficient so that the final grade reflected the weighted aspect of the task. The final grade indicated the expert opinion on how relevant the document was for the task. The inter-annotator agreement between the experts was measured by using Cohen's Kappa for two raters who provided three relevance assessments per document. Agreement was found to be substantial (Kappa = 0.684, Z = 7.04, $p < 0.001$), indicating that the expert assessments were consistent.

Additionally, we collected the position in the result list of each document returned by each participant, to better understand the reranking/scrolling tradeoff.

5.6 Data logging and data collection

For the purpose of the task completion time measurement, we recorded (1) the task duration from the start button press to the end button press. For the purpose of the effectiveness measurement, we recorded (2) bookmarked documents. After completion of both tasks in both conditions, participants were given a questionnaire to collect data on their age, gender, academic background and research experience.

We used a document set including all articles available at the Digital Library of the Association of Computing Machinery (ACM) as of the end of 2011. The information about each document consists of its title, abstract, author names, publication year, and publication venue. Articles with missing information in the metadata were excluded during the indexing phase, resulting in a database with over 320,000 documents. Both the baseline and the proposed system used the same document set and the users were presented with the top 2000 documents.

5.7 Participants

Twenty researchers in computer science (40% females) from two universities, ranging in age from 21 to 36 years old and from 1 to 8 years in research experience, volunteered to participate in the experiment. The participants were all compensated with a movie voucher that they received at the end of the experiment. All participants were assigned the same experimental tasks on both systems with systematic varying order between the systems. In this experiment, informed consent was obtained from all participants.

5.8 Apparatus

Participants performed the experiment on a desktop computer with a 27" multi-touch-enabled capacitive monitor (Dell XPS27). The computer was running Microsoft Windows 8 and both systems – being Web based – were used on a Chrome Web browser version 45.0.2454.85 m. A physical keyboard was provided for text input, whereas pointing, dragging and scrolling were performed through touch interaction. The search engine implementing the relevance estimation method was running on a virtual server and the document index was implemented as an in-memory inverted index allowing very fast response times with an average latency of less than one second.

5.9 Procedure

The tasks were described on individual instruction sheets that incorporated one of the two sets of keywords, to which we will refer as the task versions. The duration of the tasks was not constrained. To avoid introduction of confounding variables, we counterbalanced the tasks by systematically changing the order of the systems, the order of the task versions, and which task version was allocated to each system.

Considering the novelty aspect of the visualization, a training version of the tasks was devised, allowing participants to use both system with comparable proficiency. Training tasks had to be done using each system, right before the main task, using a separate set of four keywords: *creativity*, *collaboration*, *children* and *robotics*. The training started with the participant receiving a tutorial on how to use the system, then, while performing the training task, she could ask questions about either the task or the system. As soon as the training task was completed and the participant had no more questions, the participants started the actual experiment.

Participants were asked to underline the chosen answers on the instruction sheet. In the retrieval task, we asked the participants to bookmark the chosen articles. A Start/Submit button was added to both systems in the upper right corner. To be able to use each system, participants had to tap Start when ready to perform each task and Submit when they had completed it.

6. RESULTS

The results of the experiment regarding performance are shown in Table 1 and illustrated in Figure 4 with respect to the selected measures: task completion time and effectiveness, and reported according to both tasks, perception and retrieval. The mean position of selected articles in the result list is also illustrated in Figure 4. The results are discussed in detail in the following sections.

6.1 Task Completion Time

Significant differences were found between the systems in both tasks, which are discussed as follows.

6.1.1 Perception Task

The results of the perception task show that participants spent substantially less time completing the perception task when using the relevance map than when using the baseline system. The mean task duration for the relevance map was 84.23 seconds, while the mean task duration for the baseline system was 177.72 seconds. The differences between the systems were found statistically significant (Wilcoxon pair-matching ranked-sign test: $Z = 3.27$; $p < 0.001$). In conclusion, the relevance map shows 111% improvement, and was therefore more efficient for the perception task, confirming $H1$.

	Task Completion Time					Effectiveness				
	Baseline (B)		Map (M)		B vs. M	Baseline (B)		Map (M)		B vs. M
	M	SD	M	SD	Wilcoxon Test	M	SD	M	SD	Wilcoxon Test
Perception	177.72	116.20	**84.23**	39.38	**p < 0.001**	0.75	0.23	**0.89**	0.17	**p = 0.013**
Retrieval	137.53	101.66	**80.93**	70.65	**p < 0.001**	0.70	0.13	0.71	0.12	p = 0.95

Table 1: Task completion time and effectiveness results for both tasks. Task completion time is reported as a duration of the task averaged over participants. Effectiveness in the perception task is reported by mean quality of topics averaged over participants, and effectiveness in the retrieval task by mean quality of documents averaged over participants. Results showing significant improvement over the baseline are shown in bold.

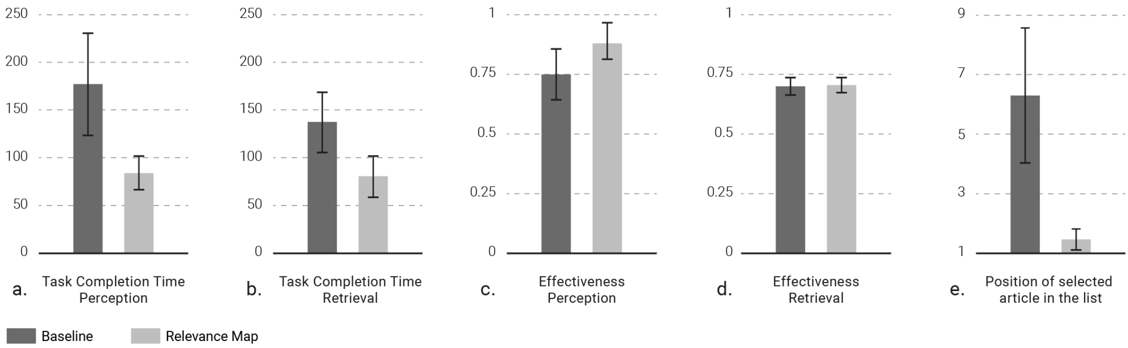


Figure 4: Results from the performance measures displayed for both systems with confidence intervals for: (a) task completion time in the perception task and (b) task completion time in the retrieval task with the mean duration (lower is better), (c) effectiveness in the perception task with the mean topic quality, and (d) effectiveness in the retrieval task with the mean document quality (higher is better). (e) Mean position in the result list of selected articles in the retrieval task.

6.1.2 Retrieval Task

In the retrieval task, participants spent substantially less time completing the task when using the relevance map than when using the baseline system. The mean task duration for the relevance map was 80.93 seconds, while the mean task duration for the baseline system was 137.53 seconds. The differences between the systems were found to be statistically significant (Wilcoxon pair-matching ranked-sign test: $Z = 3.87$; $p < 0.001$). In conclusion, relevance map shows 70% improvement and was therefore more efficient for the retrieval task, confirming $H2$.

Using the relevance map, participants selected articles close to the top in the result list, with a mean position of 1.48 (SD = 1.20), while the mean position of the selected article for the baseline system was 6.33 (SD = 7.20). The differences between the systems were found statistically significant (Wilcoxon pair-matching ranked-sign test: $Z = 4.77$; $p < 0.001$).

6.2 Effectiveness

6.2.1 Perception Task

In the perception task, the effectiveness as measured by the accuracy of the topics selected by the participants on the relevance map is 0.89, while accuracy on the baseline system is 0.75. The differences between the systems were found to be statistically significant (Wilcoxon pair-matching ranked-sign test: $Z = -2.46$; $p = 0.013$). In conclusion, relevance map was more effective for the perception task, confirming $H3$.

6.2.2 Retrieval Task

No statistically significant difference in the relevance of retrieved documents was found in the retrieval task (Wilcoxon pair-matching ranked-sign test: $Z = -0.07$ and $p = 0.95$). The fourth chart in figure 4 shows very similar results for both systems. This result fails to confirm $H4$, but it shows that the improvement in task completion time observed in the retrieval task did not impair the quality of the retrieved documents.

7. DISCUSSION

The results of the experiments show significant improvements in task completion time in both perception and retrieval, without compromising effectiveness. These results confirm hypotheses $H1$, $H2$ and $H3$.

In the perception task, participants were able to use the relevance map visualization to make decisions with greater accuracy, 111% faster. The visualization allowed the participants to understand more accurately the distribution of information with respects to the multiple aspects of the query.

In the retrieval task, documents fitting complex criteria were retrieved 70% faster using re-ranking through interaction with the relevance map. While finding documents with different relevance to several topics requires users to go through long lists of results and assess the relevance of individual documents, our proposed method for re-ranking through pointing at the map successfully narrows down the top results to documents that fit the criteria.

The quality of the task outcome was the same in both conditions in the retrieval task, which failed to confirm hypothesis $H4$. A possible reason for equal performance is the absence of strict time constraints for participants to complete the tasks. It is possible that a constrained time to complete the task would have negatively impacted the quality of the task outcome for the baseline, as the participants would not have been able to carefully examine the list to find a fitting article, but would have been forced to skim, resulting in possibly lower quality of selected topics and articles.

While our results show substantial improvements over the baseline, there is a tradeoff between the perceived simplicity of a result list and the added visual complexity of a relevance map. Interaction-wise, a result list is explored by scrolling, while a relevance map requires more complex behavior, justifying the use of a training session and tutorial. In the context of the present experiment, the necessity for a tutorial introduces a risk of influencing participants towards optimal behaviors that may outperform self-devised strategies. While our results suggest that the design of such visual interfaces can make both retrieval and perception faster, simpler interfaces may be more effective when the cost of interactions is higher, e.g. smaller devices and mobile scenarios.

We see further research directions to be addressed. First, different task complexity could be investigated and open-ended tasks explored, in which users would have more control over the search process. Second, more realistic search situations outside of our present laboratory experiment could be exploited to investigate interaction with relevance mapping and re-ranking functionality in situations in which users would have the possibility to try their own areas of interest and determine whether the suggestion effectively met their preferences and expectations.

8. CONCLUSION

Conventional systems for information retrieval are not designed to provide important insights of the data, such as relevance distribution of the results with respect to the user's query phrases. In this paper, we introduced visual re-ranking, an interactive visualization technique for multi-aspect information retrieval that helps overcome such limitations. The method proved successful in substantially improving performance over complex analytical tasks. Evaluation showed that users are able to make sense of the relevance map and take advantage of the re-ranking interaction to lower the time required to make analytic decisions or retrieve documents based on complex criteria. These results suggest that the conventional one-dimensional ranked list of results may not be enough for complex search-related tasks that go beyond simple fact finding.

9. ACKNOWLEDGMENTS

This research was partially funded by the European Commission through the FP7 Project MindSee 611570 and the Academy of Finland (Multivire, 255725, 278090 and 305739). The data used in the experiments is derived from the ACM Digital Library.

10. REFERENCES

[1] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *Proc. CHI'94*, pages 313–317. ACM, 1994.

[2] C. Ahlberg, C. Williamson, and B. Shneiderman. Dynamic queries for information exploration: An implementation and evaluation. In *Proc. CHI'92*, pages 619–626. ACM, 1992.

[3] J.-W. Ahn and P. Brusilovsky. Adaptive visualization of search results: Bringing user models to visual analytics. *Information Visualization*, 8(3):167–179, 2009.

[4] B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers. Visualizing sets and set-typed data: State-of-the-art and future challenges. In *EuroVis– State of The Art Reports*, pages 1–21. Eurographics, 2014.

[5] B. Amento, W. Hill, L. Terveen, D. Hix, and P. Ju. An empirical evaluation of user interfaces for topic management of web sites. In *Proc. CHI'99*, pages 552–559. ACM, 1999.

[6] S. Andolina, K. Klouche, J. Peltonen, M. Hoque, T. Ruotsalo, D. Cabral, A. Klami, D. Głowacka, P. Floréen, and G. Jacucci. Intentstreams: smart parallel search streams for branching exploratory search. In *Proc. IUI'15*, pages 300–305. ACM, 2015.

[7] S. Benford, D. Snowdon, C. Greenhalgh, R. Ingram, I. Knox, and C. Brown. Vr-vibe: A virtual environment for co-operative information retrieval. In *Computer Graphics Forum*, volume 14, pages 349–360. Wiley, 1995.

[8] P. P. Bonissone, R. Subbu, and J. Lizzi. Multicriteria decision making (MCDM): a framework for research and applications. *Computational Intelligence Magazine*, 4(3):48–61, 2009.

[9] K. W. Boyack, B. N. Wylie, and G. S. Davidson. Domain visualization using VxInsight® for science and technology management. *JASIST*, 53(9):764–774, 2002.

[10] M. Chalmers and P. Chitson. Bead: Explorations in information visualization. In *Proc. SIGIR'92*, pages 330–337. ACM, 1992.

[11] F. Das-Neves, E. A. Fox, and X. Yu. Connecting topics in document collections with stepping stones and pathways. In *Proc. CIKM'05*, pages 91–98. ACM, 2005.

[12] B. Fortuna, M. Grobelnik, and D. Mladenic. Visualization of text document corpus. *Informatica*, 29(4):497–502, 2005.

[13] S. Gratzl, A. Lex, N. Gehlenborg, H. Pfister, and M. Streit. Lineup: Visual analysis of multi-attribute rankings. *TVCG*, 19(12):2277–2286, Dec 2013.

[14] S. Havre, E. Hetzler, K. Perrine, E. Jurrus, and N. Miller. Interactive visualization of multiple query results. In *Information Visualization*, page 105. IEEE, 2001.

[15] M. A. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *Proc. CHI'95*, pages 59–66. ACM, 1995.

[16] E. Hetzler and A. Turner. Analysis experiences using information visualization. *IEEE CG&A*, 24(5):22–26, 2004.

[17] O. Hoeber and X. D. Yang. Interactive web information retrieval using wordbars. In *International Conference on Web Intelligence*, pages 875–882. IEEE, 2006.

[18] O. Hoeber and X. D. Yang. The visual exploration of web search results using hotmap. In *Proc. IV'06*, pages 157–165. IEEE Computer Society, 2006.

[19] C. Kang, X. Wang, Y. Chang, and B. Tseng. Learning to rank with multi-aspect relevance for vertical search. In *Proc. WSDM'12*, pages 453–462. ACM, 2012.

[20] S. Kaski, T. Honkela, K. Lagus, and T. Kohonen. WEBSOM–self-organizing maps of document collections. *Neurocomputing*, 21(1):101–117, 1998.

[21] A. Kerne, E. Koh, B. Dworaczyk, J. M. Mistrot, H. Choi, S. M. Smith, R. Graeber, D. Caruso, A. Webb, R. Hill, and J. Albea. combinformation: A mixed-initiative system for representing collections as compositions of image and text surrogates. In *Proc. JCDL'06*, pages 11–20. ACM, 2006.

[22] P. Kidwell, G. Lebanon, and W. S. Cleveland. Visualizing incomplete and partially ranked data. volume 14, pages 1356–1363. IEEE, 2008.

[23] K. Klouche, T. Ruotsalo, D. Cabral, S. Andolina, A. Bellucci, and G. Jacucci. Designing for exploratory search on touch devices. In *Proc. CHI'15*, pages 4189–4198. ACM, 2015.

[24] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *PacificVis'15*, pages 117–121. IEEE Computer Society, 2015.

[25] B. Kules and B. Shneiderman. Users can change their web search tactics: Design guidelines for categorized overviews. *IPM*, 44(2):463–484, 2008.

[26] X. Lin. Map displays for information retrieval. *JASIS*, 48(1):40–54, 1997.

[27] J. Mackinlay. Automating the design of graphical presentations of relational information. *TOG*, 5(2):110–141, 1986.

[28] J. Matejka, F. Anderson, and G. Fitzmaurice. Dynamic opacity optimization for scatter plots. In *Proc. CHI'15*, pages 2707–2710. ACM, 2015.

[29] N. E. Miller, P. C. Wong, M. Brewster, and H. Foote. Topic islands tm-a wavelet-based text visualization system. In *Proc. VIS*, pages 189–196. IEEE, 1998.

[30] E. Morse and M. Lewis. Why information retrieval visualizations sometimes fail. In *Proc. IEEE SMC'97*, volume 2, pages 1680–1685, Oct 1997.

[31] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: scalable tree comparison using focus + context with guaranteed visibility. *TOG*, 22(3):453–462, 2003.

[32] A. Nuchprayoon and R. R. Korfhage. Guido, a visual tool for retrieving documents. In *Proc. VL/HCC'94*, pages 64–71. IEEE, 1994.

[33] K. A. Olsen, R. R. Korfhage, K. M. Sochats, M. B. Spring, and J. G. Williams. Visualization of a document collection: The VIBE system. *IPM*, 29(1):69–81, 1993.

[34] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *Proc. SIGIR'98*, pages 275–281. ACM, 1998.

[35] T. Porter and T. Duff. Compositing digital images. In *Proc. SIGGRAPH'84*, pages 253–259. ACM, 1984.

[36] V. V. Raghavan and S. M. Wong. A critical analysis of vector space model for information retrieval. *JASIS*, 37(5):279–287, 1986.

[37] S. E. Robertson. Readings in information retrieval. chapter The Probability Ranking Principle in IR, pages 281–286. Morgan Kaufmann Publishers Inc., 1997.

[38] R. M. Rohrer, D. S. Ebert, and J. L. Sibert. The shape of Shakespeare: visualizing text using implicit surfaces. In *Proc. InfoVis*, pages 121–129. IEEE, 1998.

[39] T. Ruotsalo, G. Jacucci, P. Myllymäki, and S. Kaski. Interactive intent modeling: Information discovery beyond search. *Communications of the ACM*, 58(1):86–92, 2015.

[40] T. Ruotsalo, J. Peltonen, M. Eugster, D. Głowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki, G. Jacucci, et al. Directing exploratory search with interactive intent modeling. In *Proc. CIKM'13*, pages 1759–1764. ACM, 2013.

[41] J. Seo and B. Shneiderman. A rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization*, 4(2):96–113, 2005.

[42] C. Shi, W. Cui, S. Liu, P. Xu, W. Chen, and H. Qu. RankExplorer: Visualization of ranking changes in large time series data. *TVCG*, 18(12):2669–2678, 2012.

[43] B. Shneiderman. Dynamic queries for visual information seeking. *Software*, 11(6):70–77, Nov 1994.

[44] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. VL/HCC'96*, pages 336–343. IEEE, 1996.

[45] B. Shneiderman, D. Feldman, A. Rose, and X. F. Grau. Visualizing digital library search results with categorical and hierarchical axes. In *Proc. DL'00*, pages 57–66. ACM, 2000.

[46] B. Shneiderman, C. Plaisant, M. Cohen, and S. Jacobs. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, 5th edition, 2009.

[47] W. Song, Q. Yu, Z. Xu, T. Liu, S. Li, and J.-R. Wen. Multi-aspect query summarization by composite query. In *Proc. SIGIR'12*, pages 325–334. ACM, 2012.

[48] A. Spoerri. InfoCrystal: A visual tool for information retrieval & management. In *Proc. CIKM'93*, pages 11–20. ACM, 1993.

[49] J. Wise, J. J. Thomas, K. Pennock, D. Lantrip, M. Pottier, A. Schur, V. Crow, et al. Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proc. Information Visualization*, pages 51–58. IEEE, 1995.

[50] D. Xin, J. Han, H. Cheng, and X. Li. Answering top-k queries with multi-dimensional selections: The ranking cube approach. In *Proc. VLDB'06*, pages 463–474. VLDB, 2006.

[51] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *TOIS*, 22(2):179–214, 2004.

[52] J. Zhang. Tofir: A tool of facilitating information retrieval–introduce a visual retrieval model. *IPM*, 37(4):639–657, 2001.

[53] J. Zhang and R. R. Korfhage. Dare: Distance and angle retrieval environment: A tale of the two measures. *JASIS*, 50(9):779–787, 1999.

Publication IV

Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. QueryTogether: Enabling Entity-Centric Exploration in Multi-Device Collaborative Search. *Information Processing & Management*, Elsevier, Volume 54, Issue 6, 2018, pp. 1182-1202,ISSN 0306-4573, DOI: 10.1016/j.ipm.2018.04.005, May 2018.

Information Processing and Management 54 (2018) 1182–1202

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Querytogether: Enabling entity-centric exploration in multi-device collaborative search

Salvatore Andolina[*,a], Khalil Klouche[b], Tuukka Ruotsalo[b], Patrik Floréen[b], Giulio Jacucci[b]

[a] Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Finland
[b] Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland

ARTICLE INFO

Keywords:
Collaborative search
Exploratory search
Co-located search
Multi-device

ABSTRACT

Collaborative and co-located information access is becoming increasingly common. However, fairly little attention has been devoted to the design of ubiquitous computing approaches for spontaneous exploration of large information spaces enabling co-located collaboration. We investigate whether an entity-based user interface provides a solution to support co-located search on heterogeneous devices. We present the design and implementation of QueryTogether, a multi-device collaborative search tool through which entities such as people, documents, and keywords can be used to compose queries that can be shared to a public screen or specific users with easy touch enabled interaction. We conducted mixed-methods user experiments with twenty seven participants (nine groups of three people), to compare the collaborative search with QueryTogether to a baseline adopting established search and collaboration interfaces. Results show that QueryTogether led to more balanced contribution and search engagement. While the overall s-recall in search was similar, in the QueryTogether condition participants found most of the relevant results earlier in the tasks, and for more than half of the queries avoided text entry by manipulating recommended entities. The video analysis demonstrated a more consistent common ground through increased attention to the common screen, and more transitions between collaboration styles. Therefore, this provided a better fit for the spontaneity of ubiquitous scenarios. QueryTogether and the corresponding study demonstrate the importance of entity based interfaces to improve collaboration by facilitating balanced participation, flexibility of collaboration styles and social processing of search entities across conversation and devices. The findings promote a vision of collaborative search support in spontaneous and ubiquitous multi-device settings, and better linking of conversation objects to searchable entities.

1. Introduction

The impact of search on our everyday lives is unparalleled. Yet, surprisingly, search is often thought of as a solitary user activity, focusing on eliciting a user's information needs and improving search-result relevance. Recently, increasing attention has been devoted to search as a collaborative activity that is often co-located, spontaneous and initiated informally from a dialogue (Brown, McGregor, & McMillan, 2015; Morris, Fisher, & Wigdor, 2010a). Users are inspired or informed by others' searches, and can distribute search efforts, exploring the information space in parallel. Despite the increasing number of situations in which several co-located people engage in collaborative search, available devices and public screens are not effectively used for synchronous collaboration.

* Corresponding author.
E-mail address: salvatore.andolina@aalto.fi (S. Andolina).

https://doi.org/10.1016/j.ipm.2018.04.005
Received 17 July 2017; Received in revised form 24 January 2018; Accepted 10 April 2018
Available online 29 May 2018

Situations addressed by existing collaborative search systems include distributed users on mobile interfaces, or co-located interaction on tabletops. We wanted to focus on a common ubiquitous computing scenario in which several co-located users spontaneously engage in collaborative search, using personal devices as well as available large screens or projectors. In particular, we investigate what user interfaces and search systems facilitate collaboration across such devices.

Several collaborative search systems have been proposed (Table 1), supporting either distributed collaboration (Halvey, Vallet, Hannah, Feng, & Jose, 2010; Morris & Horvitz, 2007; Paul & Morris, 2009; Wiltse & Nichols, 2009), or co-located search situations (Amershi & Morris, 2008; Chung, North, Self, Chu, & Quek, 2014; Golovchinsky, Adcock, Pickens, Qvarfordt, & Back, 2008; Jetter, Gerken, Zöllner, Reiterer, & Milic-Frayling, 2011; Morris, Lombardo, & Wigdor, 2010b; Morris, Paepcke, & Winograd, 2006; Teevan, Morris, & Azenkot, 2014). Collaborative search systems have targeted different kind of devices and their typical features include support for result sharing and coordination of work. Moreover, methods and tools supporting these activities continue to be general-purpose communication systems, such as email or instant-messaging systems (Morris, 2008) or, in co-located situations, face-to-face communication. Conversely, support for exploratory search in collaborative situations has remained largely unaddressed (Hearst, 2014). We maintain that the problem of ubiquitous co-located search should be targeted as a new design problem that considers a range of devices, including smartphones, tablets, laptops, and larger public screens utilized simultaneously in the same environment. In this context, user interface concepts need to consider the opportunities provided by multitouch interaction in manipulating information directly as well as constraints, such as the limited possibility of text entry on different devices.

The main research question that we investigate is whether an entity-based user interface (Klouche et al., 2015) provides a solution to support co-located search on heterogeneous devices. Such interfaces have only been investigated in search systems tailored to individual users, especially to facilitate exploration on touch devices without a physical keyboard. How they affect collaborative search is still unknown. Entity search, recently adopted in Web-wide knowledge graphs, is an opportunity to move search from keyword-based text queries on unstructured data, toward semantic search that recognizes how text refers to different types of things, such as people, places, organizations, etc. Similarly, a user interface that allows queries and results to be formulated using entities provides information that can be manipulated in an intuitive way and might be better suited for mediated interaction between people. Recent studies on collaborative search demonstrate how it is often co-located, spontaneous and initiated informally from a dialogue (Brown et al., 2015; Morris et al., 2010a), such that search should integrate with the conversational context identifying "searchable objects." Ideally, in the future, any entity mentioned in a conversation will be searchable (Andolina et al., 2018; Brown et al., 2015; Shiga, Joho, Blanco, Trippas, & Sanderson, 2017). While the query-and-response paradigm, with long lists of document-based results, works well for individual look up search, it falls short in exploratory and mobile scenarios (Klouche et al., 2015). As in the latter cases, we posit that entity-based interfaces effectively support collaborative search in ubiquitous settings. We start by analyzing the state of the art of collaborative search situations and tools to understand current trends and needs in collaborative search and devise general design goals for our system.

We present the design and implementation of QueryTogether (Fig. 1), a collaborative search system designed for co-located exploratory search in which two or more physically co-located users search together supported by entity-centric recommendations. QueryTogether was deployed in a multi-device collaborative environment with tablets and large screens, and evaluated in a collaborative exploratory search study with nine groups, each consisting of three people. Conventional laptops and large screens, with more traditional search tools based on queries and lists were used as a baseline for comparison. The study's goal was to evaluate QueryTogether and understand how its novel design, including explicit support of exploration through entity-centric recommendations, affects collaborative search in terms of exploration support, collaboration, and engagement. The results show that, relative to the baseline, QueryTogether leads to significantly improved contribution balance and search engagement without compromising effectiveness. Interaction analyses also suggest that QueryTogether led to more effective usage of the heterogeneous devices together with improved support for diverse collaboration styles and common ground establishment.

2. Background

Collaborative search is increasingly documented as an activity that initiates spontaneously as part of co-located social interactions (Brown et al., 2015; Morris et al., 2010a). Informal opportunistic interactions among colleagues, for example, often happen by chance in common areas or cafeterias, and they may lead to conversations that end up being critical to a project's success (Isaacs, Tang, & Morris, 1996; Kraut & Streeter, 1995). As part of such conversations, people may spontaneously turn to their personal devices to collaborative search for information (Brown et al., 2015) that may or may not be familiar (Hearst, 2014). However, the systems designed to support collaborative search, until now, have focused mainly on increasing awareness of search activity and division of labor across collaborators (Morris & Horvitz, 2007; Zhang et al., 2017), leaving other important aspects of collaborative search less investigated.

Table 1 summarizes several attempts to create collaborative-search support varying from conventional distributed Web-search extensions to domain-specific tabletop designs. SearchTogether (Morris & Horvitz, 2007) facilitates remote collaboration by supporting awareness, division of labor, and persistence. ViGOR (Halvey et al., 2010) uses similar principles but in the multimedia domain. MUSE (Krishnappa, 2005) allows pairs of remote users to search for medical information. CoSense (Paul & Morris, 2009) supports remote collaborative search by focusing on sense-making and providing several rich, interactive views of users' search activities. CollabSearch (Yue, Han, He, & Jiang, 2014) is a Web search system where collaborators can save Web pages or snippets and make comments. Similarly, Coagmento (Shah, 2010) provides integrated support for communication, note-taking, and the collection of text snippets or other objects from Webpages. ResultSpace (Capra et al., 2012) provides awareness of the group activity by displaying query histories. It also supports the rating of query results and includes filtering controls based on ratings from

Table 1
Comparison of different collaborative search interfaces.

	Activity Supported	Division of Work	Sharing Mechanisms	Exploratory Support	Spatial Configuration	Devices
SearchTogether (Morris & Horvitz, 2007)	Web search	Split and multi-engine search	Query awareness and messaging	–	Distributed	Desktop
CoSense (Paul & Morris, 2009)	Web search	–	Awareness views	–	Distributed	Desktop
CoSearch (Amershi & Morris, 2008)	Web search	–	Multiple cursors on screen	–	Co-located	Desktop and mobile
WeSearch (Morris et al., 2010b)	Web search	–	Sharing results snippets	–	Co-located	Tabletop
VisPorter (Chung et al., 2014)	Text analytics	–	Push to shared display	–	Co-located	Personal device and shared screen
PlayByPlay (Wiltse & Nichols, 2009)	Web browsing	–	Send my actions	–	Distributed	Desktop and mobile
O-SNAP (Teevan et al., 2014)	Restaurant search	–	Change device orientation	–	Co-located	Mobile devices
CollabSearch (Yue et al., 2014)	Academic search/travel planning	–	Share snippets and webpages	–	Distributed	(Not specified)
Coagmento (Shah, 2010)	Web search	–	Share snippets of information	–	Distributed	Desktop
Result Space (Capra et al., 2012)	News articles search	–	Query histories	–	Distributed	Desktop
TeamSearch (Morris et al., 2006)	Image search	–	–	Faceted search	Co-located	Tabletop
Facet-Streams (Jetter et al., 2011)	Hotel search	–	–	Faceted search	Co-located	Tabletop
Cerchiamo (Golovchinsky et al., 2008)	Video retrieval (Pickens et al., 2008)	Pre-defined roles	–	Algorithmic mediation	Co-located	Shared display
Querium (Golovchinsky et al., 2012)	Scientific literature search	–	Share documents and queries	Algorithmic mediation	(Not specified)	(Not specified)
QueryTogether	Scientific literature search	–	Push entities publicly to shared display or privately to peer	Entity centric exploration	Co-located	Tablets, laptops, and large screens

Fig. 1. QueryTogether. The system enables multi-device co-located collaboration for search. It explicitly supports exploration by providing entity-based recommendations to group members. The system supports diverse collaboration styles, ranging from individual work to tight collaboration. Each group member can search privately and share interesting results or search cues either privately with individual teammates or publicly with the whole group.

individual collaborators. CoSearch (Amershi & Morris, 2008) is designed for enhancing collaborative Web search in a shared-computer setting, by leveraging multiple mice and mobile devices. The design space of using tabletops for collaborative Web search systems has been explored with several systems: TeamSearch, FourBySix Search, Cambiera, and WeSearch (Morris et al., 2010a; Morris et al., 2010b).

Two particular aspects remain poorly investigated in related research: support for collaborative exploration (Hearst, 2014), and the co-located use of heterogeneous devices. First, we review insights originating from earlier research on collaborative searches with a special focus on support for exploration. Then, we analyze how previous approaches have considered the diversity of devices and setups. Finally, we introduce the notion of entity to the context of semantic computing and delve into how it can be incorporated into novel exploratory search systems. We end by summarizing the answers to the research questions posed by this work.

2.1. Collaborative exploratory search

According to a survey on 150 users conducted by Evans and Chi (2010), most people (around 59%) engage in collaborative searches as part of an exploratory process of searching for information that may or may not be familiar. Exploratory search (Marchionini, 2006) refers to complex search tasks in which users' understanding of information needs, and the information available in the data collection, can evolve during the search session. Traditional search systems tailored for well-defined narrow search tasks may lack the necessary support for exploratory search where users can sequentially refine the expression(s) of their information needs and explore alternative search directions. Although, at the individual level, a large body of work targeting exploratory search has been produced in the last decade, surprisingly, specific features for supporting exploratory search are often been missing from collaborative search systems.

In the context of individual search, there have been many examples of techniques and systems that support exploratory information needs. Researchers in the area of semantic web have proposed several strategies to address the need for exploration in search, mainly relying on different kinds of classifications (see Wilson, Kules, schraefel, & Shneiderman, 2010 for a review). Other strategies include the use of interactive intent models that enable refining of the current search intent and identifying search directions (Ruotsalo, Jacucci, Myllymäki, & Kaski, 2014). Other approaches mainly focus on supporting exploration at the interface level, such as TweetBubble (Jain et al., 2015), which extended the Twitter interface to stimulate exploratory browsing of social media. A novel approach is the one introduced by ExplorationWall (Klouche et al., 2015), a system aimed at supporting exploration at both the interface and the system level by coupling an entity-centric recommendation engine with a novel visual interface based on parallel search streams. Unfortunately, these strategies and systems have only targeted single user scenarios.

In collaborative search situations, the main strategy for handling the need for exploration has been to rely on collaboration itself. Current systems have thus focused on enabling features to improve the collaborative process, such as awareness of other team members, communication, and division of work. However, specific features for supporting exploratory search, such as those proposed for individual users, are often missing. Although not specifically designed for exploration, TeamSearch (Morris et al., 2006) and Facet-Streams (Jetter et al., 2011) may support exploratory information needs, to some extent, by leveraging faceted search (Yee, Swearingen, Li, & Hearst, 2003). With faceted navigation, users can narrow search results by incrementally applying multiple filters, called facets (Tunkelang, 2009). Available facets (either existing categories or computed through clustering) provide an overview of the information space allowing users to choose from a list instead of coming up with a query. The exploration, however, is limited since faceted search typically relies on Boolean filters to determine a list of equally relevant items that are then sorted according to an arbitrary criterion (e.g., time, price). Moreover, such a binary approach to relevance creates limited opportunity for discovery and insights. Our approach obviates this problem by allowing the users to express preferences on entities that they find useful, presenting results sorted according to relevance estimation, without excluding information based on arbitrary criteria. Cerchiamo (Golovchinsky et al., 2008) uses a different approach by providing support for exploration in an implicit way, namely through algorithmic mediation (Pickens, Golovchinsky, Shah, Qvarfordt, & Back, 2008) and by introducing roles. Cerchiamo avoided explicit sharing and focused on ensuring that the search results of individual users would automatically be influenced by the group's search process. One limitation to this approach is that users do not have any control on what information is shared. Also, this approach only supported a form of loosely coupled collaboration, while it lacked support for tightly coupled collaboration. In a similar way, Querium (Golovchinsky, Dunnigan, & Diriye, 2012) supports exploration through system mediation. However, unlike Cerchiamo, Querium also provides communication and sharing features to enable more tightly coupled collaboration.

Our system, QueryTogether, is unique insofar as it provides support for exploration at both the interface and the system level. At the interface level, QueryTogether supports exploration through the novel representation of entities as interactive searchable objects, whereas at the system level provides smart entity recommendations for directing future searches. In this way, QueryTogether leaves control to the users who can transition between phases of loosely and tightly coupled collaboration as needed. In QueryTogether, the feature of entity-centric exploration, which has proved to be effective for individual users (Klouche et al., 2015), is adapted to a collaborative-search setup.

2.2. Multi-device search

Although collaborative search has mainly been envisioned as taking place around a single shared device (Morris et al., 2010a; Morris et al., 2010b), recent research has pointed out the importance of enabling users to initiate their searches at any time or place and with any available device based on how their information needs are triggered (Han, Yue, & He, 2015). A study conducted by Brown et al. (2015) further showed the importance of supporting mobile collaborative search at any time. Their results revealed how collaborative mobile-phone search often happens as part of informal conversations, and how searchers manage the participation of other interlocutors alongside the search itself (Brown et al., 2015).

Some collaborative systems have attempted to support multi-device interaction in search. CoSearch (Amershi & Morris, 2008), for example, supports co-located collaborative search through a shared display and multiple input devices, such as extra mice or mobile phones. Mobile phones allow individual searchers to control several functions on the shared display and to download Web pages from the shared display to their private mobile devices, thus enabling individuals to read the same Web page at their own pace. Although CoSearch provides users with more control and independence than a traditional shared interface can, the independence provided is still limited. For example, all collaborators can suggest search terms, but only the "driver" of the shared interface can execute queries.

PlayByPlay, Wiltse and Nichols (2009) supports remote collaborative search between mobile and desktop users in on-the-go scenarios by using an instant-messaging channel that allows users to send snippets of Web content. An informal study on a simulated mobile search scenario showed that the slowness of text entry was generally a problem that needed to be addressed. In the present study, however, we do not focus on remote collaboration but rather on collaborative search situations where users are physically co-located. O-SNAP, Teevan et al. (2014) proposes a multi-device mobile scenario for co-located restaurant search. The system supports the switch between individual and group search using phone orientation: vertical orientation was used for individual search, while landscape orientation was used to enter collaborative mode and have the application state of an individual (including search terms, result list, and current view) propagated across all other devices in the mode. This approach, however, posed coordination challenges when multiple users tried to control the screens at once.

In contrast to those systems, QueryTogether enables collaborative search on a wider variety of devices, including tablets, laptops and large touchscreens. This is achieved through a responsive design featuring a mechanism for typing-free querying via direct manipulation of search entities, which is particularly suitable for interaction on small-screen devices but also applies well for searching on larger screens. Unlike systems such as CoSearch, which offer limited features on smaller devices, QueryTogether provides the same set of features on all available devices. Other systems, such as VisPorter (Chung et al., 2014), support collaboration and the sharing of resources on a large variety of devices. They do not support collaborative search, but they do promote collaborative sense-making. Another difference with respect to other multi-device systems is that QueryTogether allows some devices to be used in public mode as shared workspaces, and the rest of devices in private mode for which individuals have full control over what they want to share. In this way, QueryTogether maintains the advantages of both early collaborative search systems implemented on shared tabletops and more recent systems supporting phases of independent search, such as PlayByPlay and O-SNAP.

2.3. Entity-based search interfaces

Traditional methods of information access have used a keyword-based query and a document-based response. The limited and well-defined scope of this paradigm works well in terms of addressing user intents in familiar search contexts. However, it is inefficient for complex tasks and does not allow for the efficient modeling of complex user intents in the information space (Ruotsalo et al., 2014).

Entity-based systems provide users with named entities (e.g., people, places, topics) to better address specific information needs. Conventional search engines often use Wikipedia as a comprehensive entity collection, in an entity-recommendation approach, to answer entity-centric queries. Yet, despite how comprehensive Wikipedia content may seem, most relevant information (e.g., very recent news events) is usually presented in natural language and not structured and linked according to standards.

Researchers in the fields of semantic computing and information retrieval are investigating new approaches to overcome such limitations (Balog, Meij, & de Rijke, 2010). In particular, the goal of semantic computing is to understand the naturally expressed intentions of users, determine the semantics and express them in a machine-processable format. In addition, armed with the ability to model and express users' intentions, it models and processes content and information, inferring meaning and therefore its semantics. In this way, it makes it possible to map the semantics of the user with the semantics of the content, allowing users to share meaning

(Hasida, 2007).

Recently, Web-scale knowledge graphs demonstrate through recent advances in open-information extraction (OIE) how to structure and semantically link entities from information extracted from the Web (e.g., Freebase, DBPedia, Google Knowledge Graph, Knowledge Vault (Dong et al., 2014)). Several exploratory search systems have been developed around such efforts. Some systems rely mostly on the visualization of the entity graph as provided by the database to provide links to Wikipedia articles (Nuzzolese et al., 2017). Others use interactive visualization and retrieval algorithms to rank meaningful results (Castano, Ferrara, & Montanelli, 2014; Marie, Gandon, Ribière, & Rodio, 2013).

In most cases, results are still document based, with an entity-based interactive visualization on the side. Recent entity-based systems supporting exploratory search use entities both in the results and as a query, with documents considered just another type of entity (Andolina et al., 2015; Klouche et al., 2015). Such design allows each result to be used as the start of a new search, thereby respecting the iterative and open-ended nature of exploration. When visualized as interactive objects, entities make for flexible interactions, for they can be easily moved and organized to reflect the user's state of knowledge and as a way to input new search directions. Entity-based systems have been shown to substantially reduce the need for typing, which make them especially useful for touch devices (Klouche et al., 2015). They have also been shown to foster more active behavior from users in exploratory settings, with more queries submitted and more relevant information found per unit of time as well as more branching and revisits in the search trail (Andolina et al., 2015).

3. System design

In this section, we identify design goals and describe how they were implemented in the QueryTogether system. The main theme we identified is the need to effectively support exploratory search in a collaborative and spontaneous search setting. The challenge is how to adapt exploratory features developed for single-user scenarios in a collaborative search tool by making sure to provide a good support for exploration at both the system and the interface level. Another theme that has emerged from studies is that collaborative search happens in a variety of situations and with different devices. The challenge is designing for a wide range of devices with different form factors and different input modalities. Finally, it has been noted that co-located collaborative search, although often preferred to remote collaboration, may perform less effectively (González-Ibáñez, Haseki, & Shah, 2013). In their study, González-Ibáñez et al. (2013) demonstrated that, for an exploratory search task, remotely located participants were able to explore better than co-located teams, meaning that more independence led to more diversity in information resources, which helped explore a greater variety of relevant information. Having more social presence increased interactions among the participants, but these interactions were often found to be distracting for a time-bound task. The design challenge here is how to incorporate the best of the co-located and remote approaches by making sure that people can take advantage of the more tightly coupled collaboration typical of co-located situations but also of the more independent, loosely coupled form of collaboration happening when interacting remotely.

Based on the current trends and envisioned collaborative-search practices, we hypothesized that an effective interface for co-located, collaborative exploratory search should support entity-centric exploration but also enable the flexible use of devices and support diverse working styles.

3.1. Design goals

- *Entity-centric exploration*. The units of search and collaboration can be any information entities – such as documents, keywords, and authors – that can be shared for collaboration or used as queries to trigger exploration. The design should also provide different starting points for exploration, including not only entities suggested by peers but also entities suggested by the system. System suggestions should be provided every time a new query is triggered, so users are always provided with possible directions for future exploration. In both cases the suggested entities should be encapsulated in *interactive search objects* that can be directly used to trigger new queries and explore new directions.
- *Flexible use of devices*. To study the effect of entity-centric exploration in a scenario that reflects the current trends as closely as possible, one design goal is to make the system usable from a variety of devices and thus support different modalities (e.g., mouse and touch) and different platforms. To facilitate interaction on smaller devices (and on touch screens in general), the main features such as querying and sharing should not necessarily require typing. Enabling typing-free interactions may also prevent unnecessary overhead when accepting system suggestions and thus lead to less distraction and better exploration.
- *Support for diverse working styles*. Previous work on co-located collaboration has stressed the importance of supporting a variety of working styles ranging from individual work to tight collaboration (Scott, Carpendale, & Inkpen, 2004; Tang, Tory, Po, Neumann, & Carpendale, 2006). It is important to allow for various degrees of coupling as at times, the work is more efficient if it is performed by an individual or loosely coupled. This is also important due to the fact that in some instances, it might be appropriate to allow for maintaining privacy of the information being manipulated by participants (Stefik et al., 1987). Moreover, the system should allow for flexible switching between different working styles. Users should have the option to work independently and to decide if and when to share and whether to share privately or publicly. While this could lead to more effective collaboration (González-Ibáñez et al., 2013), it could also be beneficial from a privacy perspective, as users decide what to share without needing to disclose their entire search log.

The practical implementation of the QueryTogether system consists of two parts, the user interface and the search engine, which are described in detail in the following subsections.

3.2. User interface design

The workspace is divided into two parts: the query area (Fig. 2a) at the bottom and the results area (Fig. 2b) on top. The side panel (Fig. 2f) is collapsible and always available on the right edge of the screen. It contains the reading list/share history (Fig. 2g) – where the user can visualize all saved and shared information in descending chronological order – and the user list (Fig. 2h) – where each user's name is displayed as a label along with his or her share status ("public" for shared devices and common work spaces or "private" for users with individual/private devices) (Fig. 2k).

The system uses and presents all information in the form of entities, i.e. interactive objects that embody snippets of different types of information. In the current instantiation, entities are of three types: documents, authors, and keywords. Each entity is represented by an icon, a label, and a relevance gauge and can be dragged across the workspace. This approach allows the user to make intuitive associations and to use any encountered information as a new starting point for exploration. Additional interactions include: accessing an entity's content by tapping its label and saving it by tapping its icon. Saved entities are highlighted and can be found in the reading list in the side panel.

All search queries consist of an entity or a group of entities of any type (Fig. 2e). Queries are formulated by dragging existing entities to the query area at the bottom. When entities are brought close enough together, they are considered to belong to the same group, which is made visible through a visual link. When a specific topic of interest cannot be found among existing entities, double clicking on the query area will show a textbox that allows the user to create an entity based on a custom expression.

Sharing is performed by dragging an entity to the chosen user. The recipient will instantly receive a new instance of the sent entity in his or her side panel. If the side panel is closed, a visual notification in the corner informs the user of the number of new entities received. Next to each user label, a "Message" icon (Fig. 2m) allows the user to send a short message along with an entity.

To facilitate exploration based on saved and shared entities and documents, the reading list/share history can be filtered according to a chosen collaborator, simply by tapping his or her name. Filtering based on a collaborator that uses a personal device will show only entities and documents sent to and received from that user. Filtering based on the moderator, or any collaborator using a public workspace (e.g., through a shared monitor, large screen or projector) will display the content of the common collection shared among all users. Filtering based on one's own name will display only entities that have been saved locally and ignore anything sent or received remotely.

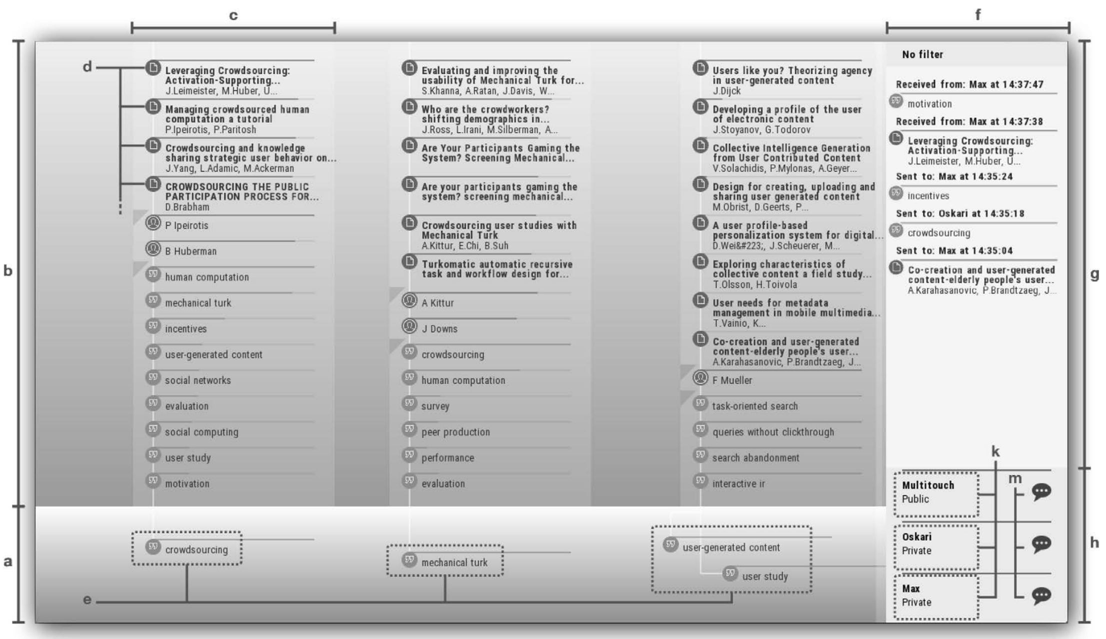


Fig. 2. The interface is composed of (a) the query area and (b) the results area. Main interface elements include: (c) search streams; (d) three types of entities: documents (brown icons), authors (red icons), and keywords (blue icons); (e) simple or composed queries; (f) the side panel, which is composed of (g) the reading list/share history and (h) the user list, with (k) a list of users and (m) a messaging option. The user composes new queries by dragging entities to the query area. A custom expressions-text input field is created by double clicking/tapping on an empty space in the query area. Entities can be saved to the reading list by dragging and dropping. The reading list also displays entities sent by collaborators. The user list shows online collaborators. Entities can be shared by dragging them to the recipient in the user list. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

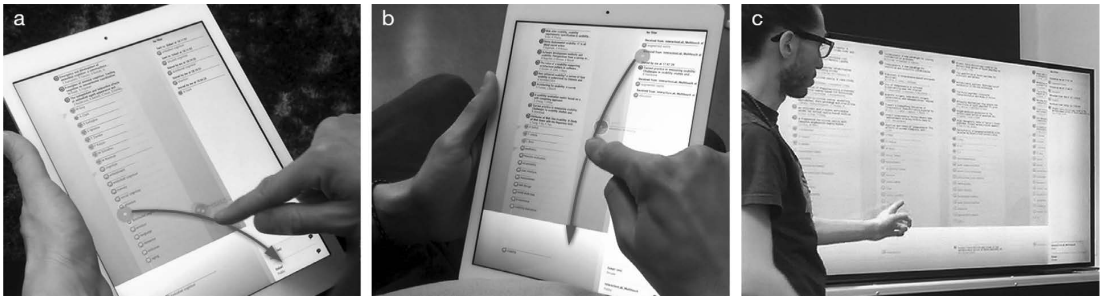


Fig. 3. This sequence illustrates the coordination and exploration process between collaborators: a) from the results of a search, User 1 chooses an entity - in this case a keyword - to send to User 2. User 1 drags the chosen entity to User 2's label in the user list. b) The received entity appears in User 2's reading list, along with information on the sender and the time it was sent. User 2 can then use it as an exploration trigger to start a new search stream by dragging the received entity to the query area. c) In the same way, entities can be sent to User 3. In this case, User 3 uses the system on a common workspace on a large screen and is set as public, so entities on the large screen are shared with all collaborators.

3.3. A walkthrough example

Max, Anna, and Oscar are three computer science students. They have teamed up to present a common project in the context of a workshop on the semantic web. They meet to look for ideas as they only have a superficial knowledge of the topic. Max and Oscar take out their tablets, while Anna takes control of the shared large multi-touch screen, assuming the role of the public user and moderator of the session.

As they each log on to the collaborative search system, their names stack up in the user list on each device. They agree to start by exploring the topic of "semantic web" at large, to find inspiration and become more familiar with related subtopics. Anna starts by tapping somewhere in the query area of the shared screen, which opens a local text box. A soft keyboard pops up, with which she types "semantic web." Based on that query, the system returns a variety of related documents and keywords organized in a vertical stream. They discuss and review the keywords suggested by the search engine that seem to be the most central. They eventually agree that Oscar will further investigate the topic "ontologies" while Max will explore "web mining." On the shared screen, Anna drags the keyword "ontologies" from the result stream to the user label "Oscar" in the user list (Fig. 3a). A new instance of the keyword appears instantly in Oscar's side panel with the mention "Shared by Anna (Moderator) at 10:36:17". Anna then does the same with the keyword "web mining" to Max. Without having to type anything, Oscar privately drags the freshly received keyword on his device toward the query area, which returns a new stream of articles and keywords all related to "ontologies" (Fig. 3b). He reads the abstracts of the retrieved articles and performs a few follow-up searches based on the related keywords. The new result sets appear as parallel streams on his interface, allowing him to compare their contents. In the same way, Max explores information related to "web mining."

As they both encounter interesting documents and keywords, they send them to the large screen by dragging them over to Anna's user label. After a little while, they decide to stop collecting new material to discuss the shared content. Anna leads the discussion on the large screen (Fig. 3c). As they review the outcome of their individual searches, they agree on which documents to keep in the list and which to dismiss. To remove an entity or a document from the list, they simply drag it out of the side panel. In the end, all three participants share the same collection of a few highly relevant documents that will make for an excellent basis to start their project.

3.4. The search engine

In order to support the document retrieval and entity oriented coordination, the search engine was designed to support two main functions: entity ranking and document ranking.

Entity ranking computes a ranking for entities given the composed query. The key idea behind the entity ranking is that the entities offered to the user by the system are based on centrality estimation given the query. For example, if the user searches for "information retrieval," she is expecting back not only entities that occur in the top ranked documents, but also ones that are central for the field of information retrieval.

Document ranking computes a ranking for the resulting documents given the composed query. Conversely to the entity ranking, document ranking is based not on centrality but solely on relevance estimation given the query.

We used a document set including more than 50 million scientific documents from the following data sources: the Web of Science, prepared by Thomson Reuters, Inc.; the Digital Library of the Association of Computing Machinery (ACM); the Digital Library of the Institute of Electrical and Electronics Engineers (IEEE); and the Digital Library of Springer. Information on each document consisted of the following: the title, abstract, author names, and publication venue.

3.4.1. Entity ranking

We represent the underlying database as an undirected entity-document graph, where each document, keyword, and author are represented as vertices, and the edges represent their occurrence in the document data.

The centrality ranking is based on the user's relevance feedback on vertices determined by dragging them into the query area. Each cluster in a query area represents a separate query that consists of a set of vertices. We use the personalized PageRank method (Jeh & Widom, 2003) to compute the ranking of the vertices. The set of vertices that the user has chosen to be part of an individual query form the personalization vector that is set to be the prior for the PageRank computation (Jeh & Widom, 2003). We compute the steady distribution by using the power iteration method with 50 iterations. The top $k = 10$ nodes from each entity category (keyword, author) are selected for presentation for the user.

3.4.2. Document ranking

The document ranking is based on a language modelling approach of information retrieval (Zhai & Lafferty, 2004), where a unigram language model is built for each document, and the maximum likelihood of the document generating the query is used to compute the ranking. We use Jelinek-Mercer smoothing to avoid zero probabilities in the estimation.

Intuitively, separating the entity ranking and document ranking approaches makes it possible to compute a limited set of entities that are likely to be the most important in the graph given the user interactions and allows users to target their feedback on a subset of the most central nodes given the interaction history of the user in any subsequent iteration. At the same time, the document ranking enables accurate and well-established methodology for ensuring relevance of the documents.

4. User study

4.1. Research questions

To demonstrate that entity-based interfaces on heterogeneous devices improve support for collaborative searches, we conducted a user study. To investigate the improvement given by QueryTogether relative to the baseline, we considered effectiveness, collaboration support, and engagement (see Sections 4.7.1, 4.7.2, 4.7.3):

1. Effectiveness: Does the QueryTogether system improve the search session effectiveness of retrieved information when compared to the baseline system?
2. Collaboration: Does the QueryTogether system enhance collaboration or lead to more balanced contribution within the group of participants when compared to the baseline system?
3. Engagement: Does the QueryTogether system lead to more engaging search behavior when compared to the baseline system?

Additionally, we aimed to investigate how and in terms of what aspects the collaboration changes, focusing on the use of searchable objects/entities, collaboration styles, and heterogeneous devices:

1. Entities as searchable objects: What is the extent of use of entities as searchable objects in conversation or through the use of QueryTogether?
2. Collaboration styles: Does the collaboration differ in tightly or loosely coupled work styles (Scott et al., 2004; Tang et al., 2006) or transitions between these?
3. Heterogeneous devices: Are there differences in how participants attend to the different devices?

4.2. Experimental design

The study followed a within-groups design with nine groups and two system conditions. Nine in-person teams of three people were assigned a collaborative search task. The system conditions were the full QueryTogether and a baseline version of the system that did not have the design features of QueryTogether. Each group performed two tasks: one with the support of QueryTogether and one with the baseline. The conditions and tasks were counterbalanced by changing the order in which the two tasks were performed and the order in which the groups were subjected to each condition.

4.3. Baseline

As the number of factors affecting the design did not allow for baselines where each factor would be studied in isolation, we designed a baseline system based on the following rationale:

- *De-facto practice.* As the study did not focus on a single isolated system factor but rather a whole system design, we wanted to compare QueryTogether with a set of tools representing an authentic *de facto* work practice. A recent survey study (Morris, 2013) suggested that despite the increasing availability of tools designed specifically to support collaborative search scenarios, users may be reluctant to adopt them. The de-facto practice of performing collaborative information seeking and search is based on re-purposing everyday communication technologies, such as online document sharing combined with conventional information search systems.
- *All other system features being equal.* As the study compared a system with the proposed design features, and a setup where the participants were using de-facto practice, we wanted to ensure that all other system features were as equal as possible. These

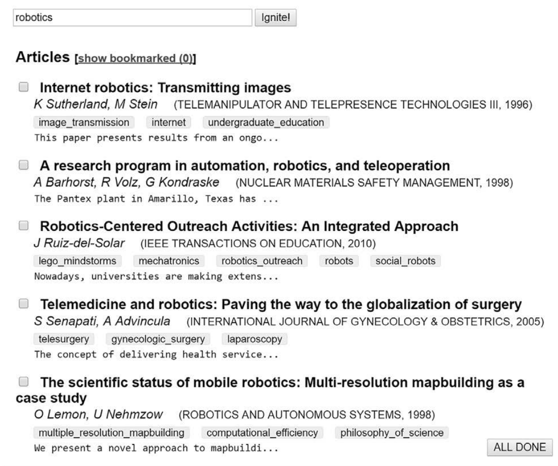


Fig. 4. A screenshot of the baseline system. It uses the same underlying dataset and ranking model as the QueryTogether system. The interaction with the system is limited to typed queries and the result presentation relies on a ranked list. Papers can be bookmarked to build a personal list. Such a list is hidden by default but can be expanded by clicking on the link "show bookmarked".

include the database and the indexing and ranking functions of the search engine. In other words, we used the exact same underlying database, the search engines used the same state-of-the-art ranking function, and we indexed the data using the same inverted indexing technique. This ensured that none of these would become confounding factors in the study.

- *Interface mimicing.* We wanted to mimic the interface design of publicly available Web search engines and digital library search systems with which the participants would be familiar. These interfaces typically display a search box and a list of query results.

The baseline, illustrated in Fig. 4, was implemented based on this rationale. The baseline was a control condition that mimicked the conventional search interfaces, allowed for isolating confounding factors related to data or search engine functionality, and enabled the users' de-facto collaboration practice to be conveyed by combining the baseline system with the tools that the participants would use in real-life situations.

We selected Google Docs as the information gathering and sharing platform, as this is the most commonly used collaborative platform available and the participants were likely to be familiar with the platform.

4.4. Procedure

The experiments took place in a dedicated interaction laboratory, shown in Fig. 5. The participants in the group were first debriefed on the experimental procedure and the purpose of the study. Then, they signed an informed consent form to take part in the experiment. The participants were also told that they were free to withdraw from the experiment at any moment with no consequence. The experimenter then illustrated the system, the interface functionalities, the devices, and the tools that the participants

Fig. 5. The experimental setting with the two experimental conditions. In the QueryTogether condition (left), two of the participants were equipped with tablet computers, and the moderator was equipped with a laptop computer. The QueryTogether was used in all of the computers. In the baseline condition (right), all the participants used laptop computers. The moderator screen was projected to the large screen in both conditions. The figures are captured frames from a video recording of an actual experimental session.

could use. Then, the participants were asked to try out an exemplary task. This ensured that the participants were aware of how to operate the system and use the tools and devices as intended. Prior to starting the actual session, each group first had to elect a moderator who would have the responsibility of leading the collaboration and managing the final outcome. After this, the experimented made sure that the participants agreed that the system functioning, the tasks, and the roles of the participants were clear, and the actual experimental session began.

Two of the participants sat in chairs with an optional table, and the moderator sat at a separate desk. The three participants were placed in a triangle facing each other at a distance of approximately 2 m to make it easy to see each other and communicate. When using the baseline, all participants used laptop computers. When using the QueryTogether system, two participants used tablets (9" diagonal screen size), and the moderator used a laptop computer. In both conditions, the moderator's computer was mirrored onto a large screen that faced the other two participants.

The participants had 20 min to complete the task with each system. After completing the task, participants were asked to fill out the User Engagement Scale questionnaire. The entire experimental session lasted about 80 min. Each participant received two movie tickets as compensation for participating.

4.5. Participants

Altogether, 27 people (8 women) were recruited to take part in the experiments. Participants were recruited from the computer science department of two Universities, and they all had at least some research experience. The participants were assigned to one of nine groups of three people each. Participants in the same group knew each other. Recruiting was done via word-of-mouth and specialized mailing lists. To reduce the effect of social dynamics during group formation, which could distract from the main task (Tuckman, 1965), and to study a more realistic situation, the recruiting message specified that participants were supposed to know each other, thus requiring to sign up in groups of three.

The mean age was $M = 29.30$ ($SD = 3.71$) and the levels of education were: 26% PhD, 52% master's degree, 11% bachelor's degree, 11% high school. When asked to self-assess their ability to successfully find information using a Web search engine, 63% chose "average" and 37% chose "expert."

4.6. Tasks

The task was created to support an exploratory search scenario. To ensure that the participants had at least some research experience while not being experts on the proposed topics, we used a screening questionnaire. The task was formulated as follows: *Your group has been asked to write a scientific review on topic X. As part of the task, you have to perform a bibliographic search on topic X. You have 20 min to find at least 20 relevant papers that cover as many subtopics as possible.* Two topics were used in the evaluation sessions: "crowdsourcing" and the "semantic web."

Participants in the QueryTogether condition collected the relevant papers in the reading list (Fig 2g), while participants in the baseline condition used a shared Google doc. All the sessions lasted 20 min. Although participants could collect as many relevant papers as needed, they were instructed to keep only the 20 most relevant papers found and discard those that exceeded that quota.

4.7. Data collection and analysis

The data collected during the study were based on observations, questionnaires, logged data, and interviews. The entire trial was video-recorded to allow offline analysis. Two researchers reviewed the recordings of the session independently, transcribing the utterances that took place. The transcriptions were subsequently coded in terms of types of utterances and collaboration strategies. The same researchers then went through a second cycle of video analysis to quantify the usage of the large public screen. Transcription and coding were performed using the software ATLAS.ti. Among other things, the software enabled associating timestamps to different transcriptions and codes.

In order to operationalize the research questions, a set of actionable measures were defined to quantify each of the aspects: effectiveness, collaboration, and engagement. In particular, we analyzed collaboration based on contribution balance, usage of the public screen, and the quantity and type of discourse produced while solving the task. Qualitative analysis of the video recordings were also performed in order to better understand the patterns of collaboration that took place.

4.7.1. Effectiveness

In order to measure the quality of the information retrieved, we quantified the effectiveness of the search session. Since the baseline returns lists of documents while QueryTogether returns lists of mixed-type entities, we chose to solely measure the quality of the retrieved documents. Our main effectiveness measure is an adapted version of *S-Recall* (Zhai, Cohen, & Lafferty, 2003) for the search task outcome, as the task was to find at least 20 relevant documents that would cover as many subtopics as possible. The subtopic retrieval measure reflects the goal of exploration, for it measures the coverage of many different subtopics rather than merely relevance. This means that the usefulness of a document depends on other documents and their topical distribution (as determined during the search session.)

The original formulation of S-recall reflects the proportion of unique subtopics covered by a ranked list. Such a definition is meant for single queries and does not fit well the case of search sessions based on multiple iterations of queries. We thus adapted the S-Recall measure to reflect the proportion of relevant and unique subtopics covered by a set of documents. As an approximation of the subtopics covered by a given document, we used the author-defined keywords of that document that were available as metadata in our dataset. S-Recall was computed over time in order to get a sense of how fast users were able to reach a good coverage of relevant information across different subtopics.

Additionally, we used more traditional information retrieval metrics such as precision, recall, and F-measure. All the measures were calculated at the group level to quantify the overall quality of the information retrieved by the group during the search session. The measures were computed using a pool of documents constructed from the system logs. This ensured that the pool contained all documents found by any of the groups. Domain experts assessed the relevance of each retrieved document on a binary scale: relevant or irrelevant. Author-defined keywords for each relevant document were considered relevant subtopics.

4.7.2. Engagement

The degree of user engagement is a strong indicator of search performance (White & Roth, 2009). For example, the extent to which the user is focused on the task can indicate whether the system is fulfilling its role in supporting search activities (White & Roth, 2009). To evaluate the engagement, we used the User Engagement Scale (UES) for exploratory search (O'Brien & Toms, 2013). The UES questionnaire in its original form includes 31 questions in six different dimensions: Aesthetics (AE), Focused Attention (FA), Felt Involvement (FI), Perceived Usability (PUs), Novelty (NO), and Endurability (EN) aspects of the experience. Here, we use the revised form of UES comprising 28 questions and four factors. With respect to the original version, while FA, AE, and PUs remained distinct factors, items from the NO, FI, and EN subscales were joined to form one factor (O'Brien & Toms, 2013). In accordance with (O'Brien & Toms, 2013), for each question in UES, participants indicated the extent to which they agreed with each statement about their exploratory search experience on a 7-point Likert scale from strongly disagree (1) to strongly agree (7).

4.7.3. Contribution balance

The balance of contribution between group members gives an indication of the quality of the collaboration in both conditions, as groups for which all the relevant subtopics were found by only one or two members may be considered to be collaborating less than those groups in which each member found a similar number of relevant subtopics. Previous studies (Salomon & Globerson, 1989) revealed that unbalanced contributions lead to a loss of motivation in team members, which further leads to the loss of productivity. The reasons for this include the several negative social dynamics happening in teams with unbalanced contributions. In such teams, for example, the less hard-working members may suffer from the *free-rider* effect, deciding that their efforts are dispensable, whereas the hardest-working members, who do most of the work for the whole team, may gradually decide to expend less mental effort to avoid the feeling of being taken advantage of, the so-called *sucker* effect (Salomon & Globerson, 1989).

We considered the number of unique and relevant subtopics each group member contributed and used the Gini coefficient of inequality to measure the balance of contribution. This measure is often used to measure inequality of income distribution (Firebaugh, 1999), but it has also been used to measure participation across group members using an interface (Rogers, kyung Lim, PhD, & Marshall, 2009). It is a ratio measure that can be used to compare inequality across cases with different overall measures. The Gini coefficient ranges from 0 (*no inequality*) to 1 (*total inequality*). In our case, smaller values of the coefficient would indicate a more equitable distribution of unique and relevant subtopics.

4.7.4. Public screen usage

The attention to the common public screen was also considered an the indicator of the quality of collaboration. In particular, this measure reveals how participants took advantage of heterogeneous devices available. Video recordings were analyzed by two independent researchers who counted the number of glances toward the public screen by those participants using private devices. The participant acting as moderator was excluded from this calculation as (s)he was working directly on the public screen. One session from the baseline and one from QueryTogether were randomly selected to be coded by both raters to assess the interrater reliability. Cohen's kappa score was substantial (0.72).

4.7.5. Usage of entities in conversation and QueryTogether

The usage of entities in QueryTogether was derived from system logs. To understand the usage of entities in the conversation, the type of discourse produced while carrying out the task was analyzed. For this purpose, we developed a coding scheme that included six kinds of communicative acts used during the task: sharing (S), sharing prior knowledge (PKS), clarification questions (CQ), queries (Q), answers (A), and reports (R). In particular, sharing was further qualified based on the specific entity it regarded (documents: D, authors: A, or keywords: K). We defined sharing as utterances that contained an entity that could be used as searchable object by other collaborators. For example, "I think Mechanical Turk is a good keyword" would be coded with S(K), meaning a sharing of a "keyword" entity. For the S code, we mainly considered the cases in which the entity was mentioned for the first time in the session. The cases in which the entity was not yet acknowledged by other collaborators but was mentioned for a second time by the same person were also coded with S. The cases in which it was clear that the sharing derived from prior knowledge were coded with PKS, for example, "I know Tim Berners-Lee is the father of Internet." We defined clarification questions as those utterances asking for more information about an entity, for example, "What's Mechanical Turk?" Queries were defined as generic questions about an entity, such as, "Did anyone use collaborative computing (a previously shared keyword)?" Answers were defined as replies to clarification questions or queries. Reports were defined as coordinated utterances reporting actions done with entities, such as, "Don't do collaborative computing. I did that already" or "I'm already looking at social computing." One session from the baseline and one from the QueryTogether condition were randomly selected to be coded by both raters to assess the inter-rater reliability. Cohen's kappa score was substantial (0.77).

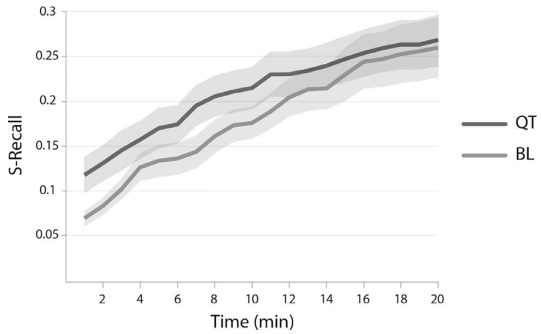


Fig. 6. S-Recall over time (averaged across groups).

4.7.6. Collaboration styles

The degree to which collaborators work closely together or independently is referred to as tight versus loose coupling (Dewan & Choudhard, 1991). To understand the styles of collaboration in QueryTogether and the baseline, we measured the intervals of time spent in *tightly coupled collaboration* and in *loosely coupled collaboration*. An interval of time was coded as tightly coupled collaboration (TC) when group members engaged in intense verbal communication. Conversely, the intervals of time characterized by absence of communication or one-way communication from one group member to the others were assigned the code referring to loosely coupled collaboration (LC). One session from the baseline and one from the QueryTogether condition were randomly selected to be coded by both raters to assess the inter-rater reliability. Cohen's kappa score was substantial (0.66). The coded transcriptions were also used to measure the *support to flexible coupling* by counting the number of transitions between the two modalities.

5. Findings

The findings from our evaluation can be grouped by three themes: effectiveness, user engagement, and collaboration. Qualitative analyses were also carried out to examine in more detail the patterns of collaboration and interaction that took place.

5.1. Effectiveness

The average number of queries per session was 32.89 ($SD = 11.62$) for QueryTogether, and 28.89 ($SD = 12.70$) for the baseline. Fig. 6 shows the S-recall for both system conditions averaged over groups for both system conditions. The users find approximately equal subtopic coverage for the search session using both systems. However, the results show how users in the QueryTogether condition were able to cover a large part of relevant subtopics in the earlier phase of the search session than users in the baseline condition.

The precision, recall, and f-measure values at group level for QueryTogether were $M = 0.45$ ($SD = 0.20$), $M = 0.19$ ($SD = 0.07$), and $M = 0.24$ ($SD = 0.07$), respectively, and the same values for the baseline were $M = 0.49$ ($SD = 0.32$), $M = 0.21$ ($SD = 0.06$), and $M = 0.25$ ($SD = 0.06$), respectively. Paired t-tests showed no significant differences among the two systems in effectiveness.

5.2. User engagement

Table 2 shows the results from the UES questionnaires. A Wilcoxon signed rank test with Holm-Bonferroni correction indicated that QueryTogether UES scores were statistically significantly higher than the baseline. The results suggest that user engagement was improved in the QueryTogether condition.

5.3. Contribution balance

The per-group Gini coefficients of the number of unique and relevant subtopics each group member contributed were $M = 0.33$ ($SD = 0.17$) for QueryTogether and $M = 0.70$ ($SD = 0.14$) for the baseline. A paired t-test shows that the difference is significant ($t(8) = 5.36$, $p < .01$), which indicates that a more balanced participation was observed for QueryTogether than for the baseline (Fig. 7).

Table 2
Results from UES Questionnaire. A Wilcoxon signed rank test with Holm-Bonferroni correction indicated that QueryTogether UES scores were statistically significantly higher than the baseline.

	QueryTogether			Baseline			Comparison
	M	SD	Mdn	M	SD	Mdn	Wilcox. Test
PERCEIVED USABILITY							
I felt discouraged while using this system	2.70	1.44	2	4.22	1.80	5	
I felt frustrated while using this system	2.89	1.53	2	4.30	1.88	5	
I felt annoyed with using this system	2.81	1.61	3	4.07	1.90	5	
This search experience did not work out the way I had planned	3.56	1.67	4	4.59	1.60	5	$Z = 2.10$
I could not do some of the things I needed to do using this system	4.22	1.80	5	4.74	1.65	5	***p* = .04**
I found this system confusing to use	2.85	1.43	2	3.04	1.74	3	
Using this system was mentally taxing	2.70	1.38	2	3.41	2.02	3	
This search experience was demanding	3.81	1.30	4	3.93	1.69	4	
I felt in control of the searching experience	4.44	1.22	5	3.93	1.38	4	
NOVELTY, FELT INVOLVEMENT, ENDURABILITY							
I felt interested in my searching tasks	4.89	1.28	5	3.41	1.50	3	
The content of this system incited my curiosity	5.22	1.12	5	3.15	1.43	3	
My search experience was fun	5.33	1.14	5	2.59	1.22	3	
I felt involved in the searching tasks	5.19	1.47	5	3.48	1.72	3	$Z = 4.00$
My search experience was rewarding	4.44	1.22	4	3.19	1.30	3	***p* = .0002**
I would recommend this system to my friends and family	4.30	1.32	5	2.37	1.24	2	
I was really drawn into my searching tasks	4.22	1.15	4	2.89	1.31	3	
I consider my search experience a success	4.19	1.59	5	3.19	1.33	3	
Searching using this system was worthwhile	4.85	1.43	5	2.89	1.22	3	
AESTHETIC APPEAL							
The screen layout of this system appealed to my visual senses	5.63	1.11	6	2.67	1.47	2	
The system interface is aesthetically appealing	5.67	1.24	6	2.52	1.31	3	$Z = 4.31$
The system interface is attractive	5.48	1.22	6	2.15	1.20	2	***p* = .0001**
I liked the graphics and images used by this system	5.59	1.28	6	2.33	1.30	2	
FOCUSED ATTENTION							
I was so involved in my searching task that I lost track of time	3.59	1.55	3	2.78	1.48	3	
The time I spent searching just slipped away	3.89	1.58	4	2.78	1.60	3	$Z = 2.41$
I lost myself in this searching experience	3.26	1.53	3	2.37	1.28	2	***p* = .03**
I blocked out things around me when I was using this system	3.52	1.55	4	2.48	1.53	2	
I was absorbed in my searching task	4.11	1.48	4	3.11	1.60	3	

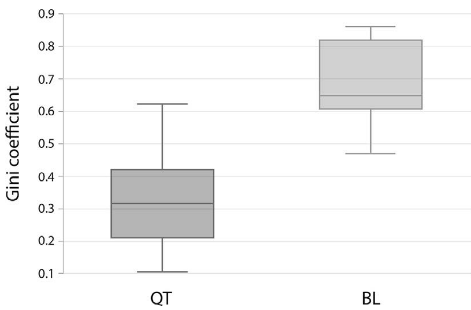


Fig. 7. Gini coefficient of the distribution of unique and relevant subtopics found by individual group members within a group.

5.4. Interaction analyses

5.4.1. Entity usage

The average number of shared entities was 30.44 ($SD = 8.05$), 3.22 of which ($SD = 2.44$) were search cues (keywords or authors). In the QueryTogether condition, an average of 32.33 search terms were used. In only 47% of the cases ($M = 15.22$, $SD = 8.30$) did users rely on the keyboard to generate new query terms. In the remaining 53% of the cases, query terms were generated by dragging a system or peer suggested entity to the dock. More specifically in 46% of the cases ($M = 15.00$, $SD = 6.84$), the used search terms were suggested by the system, while peer suggestions accounted for 7% of the cases ($M = 2.11$, $SD = 1.83$).

Fig. 8 shows how entities were used in the conversation that took place while carrying out the task. While the sharing of entities was almost the same for QueryTogether and baseline, in the QueryTogether condition, we observed slightly more usage of entities in

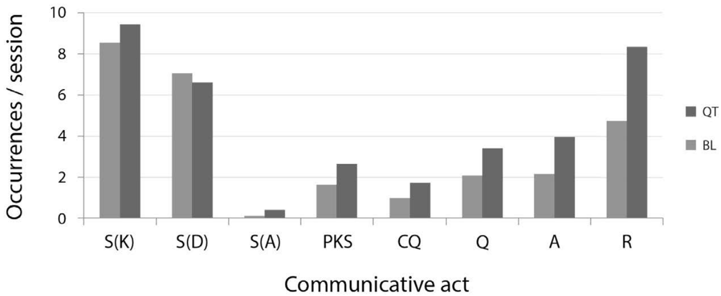


Fig. 8. The usage of entities in the conversation that took place while carrying out the task. In the X axis the different kind of communicative acts coded: S(x) = sharing, $x \in$ {K = keyword, D = document, A = author}, PKS = prior knowledge sharing, CQ = clarification questions, Q = queries, A = answers, R = reports. In the Y axis the average frequency of those codes in a session.

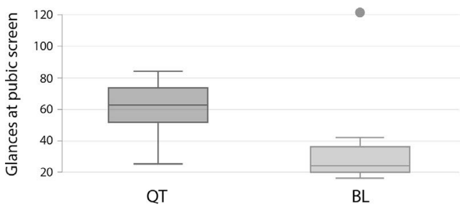


Fig. 9. Number of glances at the large public screen.

the context of sharing prior knowledge, clarification questions, general queries, answers, and reports. The low frequency of the event S(A) also shows the scarce role played by the entity "author" in the study.

5.4.2. Public screen usage

In QueryTogether participants looked at the large public screen 60 times on average (SD = 18.32), while in the baseline the number of glances at the public screen was M = 36 (SD = 33.28). An inspection of Fig. 9 shows that the number of glances at the public screen was generally higher in the QueryTogether condition.

5.4.3. Collaboration styles

The time spent in tightly coupled collaboration was M = 13.29, SD = 5.13 min in QueryTogether, and it was M = 11.49, SD = 7.18 min at the baseline. The number of transitions between tightly coupled and loosely coupled collaboration was M = 10.11 (SD = 4.78) for QueryTogether and M = 6.33 (SD = 4.72) for the baseline. This result suggests that QueryTogether supported a more flexible coupling. By inspecting Fig. 10 is it possible to get a qualitative overview of how different styles of collaboration took place in both conditions.

5.4.4. Rate of verbal communication

The mean number of utterances per minute was calculated by running a matlab script over the timestamped transcriptions. Fig. 11 shows the change over time in the rate of dialogue. The chart shows a similar amount of dialogue production in the first part of the session between the two conditions, while more dialogue is produced by participants in the QueryTogether condition during the second part of the session. This is compatible with what is suggested by Fig. 10, which showed that, compared to the baseline, the second part of the session for QueryTogether was characterized by more tightly coupled collaboration, thus having more intense verbal communication.

5.4.5. Collaboration and coordination strategies

Fig. 10 provides a general overview of how different styles of collaboration took place in both conditions. In this section we start from qualitative analyses of video recordings and logs of the sessions to further understand the patterns of collaboration and coordination strategies that emerged during the study. The way the groups initiated and managed their participation was found to vary across sessions regardless of the condition.

Groups initiated their collaboration, either by working on their own or establishing common ground based on initial searches and prior knowledge, or by dividing the work among team members. The differences across conditions regard the way in which those strategies were implemented. For example, while a similar amount of entities were mentioned as part of the collaborative task

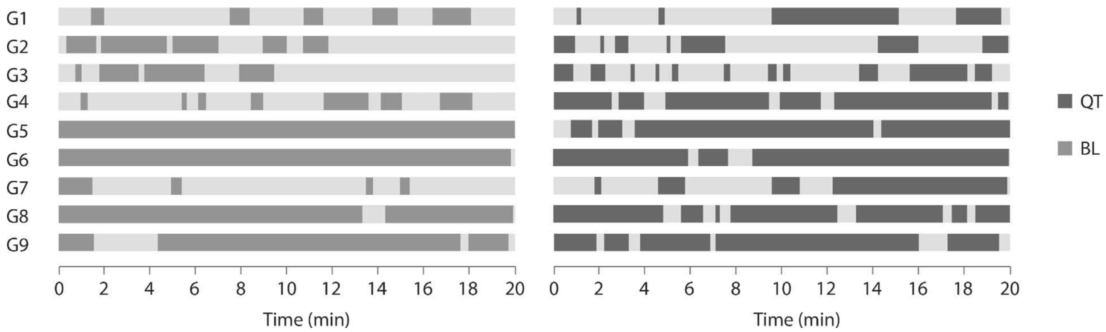


Fig. 10. The way in which groups collaborated in QueryTogether (QT) and the baseline (BL). Colored intervals (red for QT and blue for BL) indicate tightly coupled collaboration, in which group members engage in intense verbal communication. Light gray indicates loosely coupled collaboration where the verbal communication is either absent or one way from one group member to the other. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

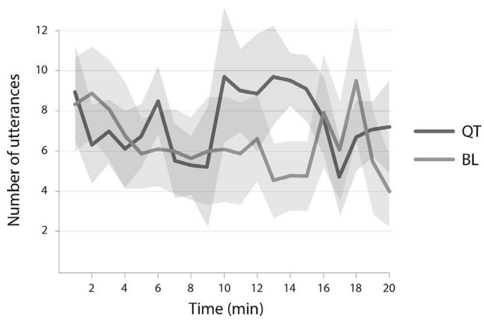


Fig. 11. Utterances per minute (averaged across groups).

(Fig. 8), in QueryTogether, the associated *searchable objects were readily available in the system* and could be used for search. Additionally, QueryTogether afforded strategies not easily implemented with the baseline. Below we use exemplary excerpts of conversations as the starting point for discussing those strategies.

In the QueryTogether condition, three groups initiated their collaboration with the *labor division*, compared to one case of the baseline. An example of labor division in QueryTogether was where participant P10 suggested to split the search effort and shared keywords with the other group members to facilitate the process:

Fragment 1	*Transcript 1*
P10 (Mod):	So: let's share some of the keywords?
	(3.9)
P12:	I would try with just crowdsourcing in general =
P10 (Mod):	= Yes, but I meant, we can combine those keywords (.)
	I can take crowdsourcing with social networks, you take crowdsourcing with user studies
P12:	<u>Ok</u>
P10 (Mod):	So we can explore together
P11:	<u>Alright</u>

This fragment exemplifies the importance of the system suggestions in providing directions for exploration. The entity sharing mechanism also allowed every group member to easily take control over a specific search direction. On the other hand, in the baseline condition, situations such as these relied mostly on verbal communication, and this caused some difficulties. For example, when participant P5 suggested to P6 to investigate "reCaptcha", P6 replied by asking to spell the keyword, which was an operation that took time and might have been distracting for other group members. In the baseline condition, the only participants able to more effectively share information were the moderators, as their search process was shown on the large public screen. The interaction, however, was still limited. As an example, participant P23 pointed at the large screen and asked the moderator to expand a particular abstract. In this situation, QueryTogether would have allowed the sharing of the actual entity, allowing P23 to take control over the document without needing to ask the moderator to perform actions on his behalf. Situations like these, seem to confirm the intuition gained from Fig. 9, in that the public screen could be used more effectively in QueryTogether. This also provides an example of unbalanced contribution in the traditional collaborative setting, with one person doing most of the work while the others provide guidance.

Another interesting example of collaboration that was supported by the searchable objects provided in QueryTogether regards the common ground establishment (Clark, Brennan, et al., 1991). An example is when P24 shared some prior knowledge and performed a *search cue suggestion*.

Fragment 2	*Transcript 2*
P24:	Do you know about this
	(2.5)
	I suppose you have heard about the Mechanical Turk kind of things?
P22 (Mod):	How does that work?
	((P24 explains))
	((P23 pushes the Mechanical Turk keyword to the public display))
P23:	I sent it
P22 (Mod):	So, alright (.) Mechanical Turk
	((P22 drags the shared keyword to the query area and triggers a new query))

In this fragment, P24 started a discussion about a relevant keyword in the domain in order to suggest using that keyword for the search. This resulted in a keyword sharing and a query triggered from the shared keyword. The keyword suggestion was used to facilitate discussion and establish a common ground among participants. It is also very interesting to note that the entity was actually sent to P22 by P23 and not by P24. Before P24 could finish his reasoning and share the entity, P23 had already suggested it from the system and could use it, in this case, for sharing it with P22. This shows the double role of entity recommendation in QueryTogether. One aspect is that it represents important triggers for discussion. The other aspect is that the system provides searchable objects for what could likely become a matter of discussion.

However, it is also worth noting that participants were sometimes so involved in their search task that they ignored the suggestion or delayed its acknowledgement. For example, when participant P01 suggested "Tim Berners-Lee" was the father of Semantic Web and shared the related entity, other group members initially ignored that suggestion; after about 7 min, P03 finally asked for clarification on the role of the suggested author, and discussion on the topic got started. This example supports the intuition that people in the QueryTogether condition were very involved in the search task during the first part of the session. They were often able to reach a good level of productivity soon without necessarily resorting to tightly coupled collaboration, and they spent more time in intense discussion only during the second part of the session, after having contributed consistently to the final outcome. Nevertheless, the example above shows how QueryTogether enabled the flexible change of collaboration style. When entities were not acknowledged soon in the baseline, they would be easily forgotten. In QueryTogether, however, participants could associate the verbal sharing of an entity with the sharing of the corresponding searchable object. The entity object would then stay visible in the share history (Fig. 2g). That enabled easy recall, allowing persons who were temporally working in an individual mode to finish their task, acknowledge the previously shared word, and switch to a tightly coupled collaboration form with other members of the group. The example above illustrates a peculiar pattern of collaboration of QueryTogether not being easily implemented in the baseline that we call *leave a note*. Video analyses show six occurrences of the pattern in QueryTogether, compared to one occurrence in the baseline in the entire study. This pattern could have partially contributed to the more flexible coupling observed in QueryTogether (cfr Section 5.4.3).

6. Discussion

The starting point of this research was that co-located collaborative search is increasingly observed in spontaneous situations (Brown et al., 2015; Morris et al., 2010a), while previous work mostly addressed bespoken arrangements such as tabletops, or distributed setups with desktop and mobile applications. Moreover, previous research highlighted the need to more explicitly support exploration in collaborative search situations (Hearst, 2014). The present work investigated whether an entity-based interface that successfully supports individuals in an exploratory search provides the right basis for designing for co-located collaborative search across personal devices and public screens. We developed goals for designing such a solution, particularly entity-centric design as well as support for a diversity of devices and working styles. The resulting system, QueryTogether, provides support for exploration at both the interface and the system level by combining a novel representation of entities as interactive searchable objects and smart entity recommendations. A study with 27 participants (nine groups) provided important insights on how our design improved collaborative search, while also revealing key challenges for future developments of the paradigm.

Contribution balance. Our study showed how our system design affected the way in which co-located groups collaborate in search. A main finding was the substantial difference between the Gini coefficients between the conditions, suggesting that QueryTogether led to more balanced participation in terms of contributions to the final outcome when compared with the baseline. A possible explanation is that engaging in an exploratory search task requires increased user effort as the topic, and, consequently, the key search terms are unknown at the beginning of the task. Evaluation apprehension, or similar negative social dynamics, may inhibit the more reticent participants to ask for help. In the baseline condition, many people limited their search to the obvious queries without being able to refine them. In contrast, in QueryTogether, the participants were able to avoid the typically observed situation in which one user searches while others watch and provide guidance. The explicit support for exploration, via system or peer suggestion of entities, always provided participants with possible directions for exploring an unfamiliar topic, thus making them more active. In particular, participants mostly used entities suggested from the system, whereas peer suggestions were used less. Prior research on remote collaboration has highlighted how participants often don't make heavy usage of specific coordination features involving peer recommendations (Morris & Horvitz, 2007). It is not surprising then that the same applies to co-located situations, where people have the possibility to coordinate by simply talking to each other. Our findings suggest that supporting system suggestions instead seems to be a more promising choice for the design of interfaces that foster equitable contributions in co-located collaborative search. Our study also suggests that the balance of contributions among team members may be associated with user engagement, which was found to be higher in the QueryTogether condition. This may have contributed to making participants less prone to social loafing, which is a phenomenon observed in individuals who rely on the group to exert less efforts (Williams & Karau, 1991).

Multi-device collaboration. Our results suggest that our design could be effective in a multi-device setting. In this paper, we envision a future in which people can start searching together at any time and with any available devices, including smartphones, tablet, and pervasive public-displays. While a significant portion of searches nowadays are executed on mobile phones, some complex tasks that require several iterations, such as those involving exploratory search of scientific literature, are still mainly performed on laptops. In the QueryTogether condition of our study, two of the three participants used tablet devices to perform exploratory search tasks. Although these kinds of tasks may seem easier on a desktop computer due to the intrinsic limitations of tablets smaller screen sizes (see Miller, Sumeeth, & Singh, 2011), as our measure of S-recall over time shows, our design allowed participants to be as effective with tablets as with laptops and more traditional search tools. Moreover, our analyses of effectiveness and verbal activity over time revealed that participants in the QueryTogether condition found most of the relevant subtopics sooner compared with participants who used the baseline, thus allowing more time for discussion and the establishment of common ground (Clark et al., 1991; Hertzum, 2008).

Entities as interactive search objects. Visualizing entities as interactive search objects seemed to have played an important role in supporting collaborative search on touch devices. We observed a reduced need for typing, with 53% of search terms created by dragging and dropping entities into the query area. Typing on soft keyboards is slower and more tedious than typing on physical ones (Hartmann, Morris, Benko, & Wilson, 2009; Hinrichs, Hancock, Carpendale, & Collins, 2007). This can negatively affect the efficacy and user experience of search applications, which are often heavily based on text entry activities. Our results are in line with findings from prior research (Klouche et al., 2015) indicating that, when given the opportunity, people prefer to directly manipulate search entities than type search terms on soft keyboards.

Flexible collaboration style. Our qualitative findings revealed some interesting details on how the collaboration took place. The analysis of the collaboration styles suggests that QueryTogether supported a more flexible coupling (Dewan & Choudhard, 1991), meaning that the system made it easy to switch between phases of individual work with little coordination, and phases of tightly coupled collaboration, in which participants engaged in intense verbal communication. Notably, the analyses revealed a peculiar strategy which seemed to be particularly suited for this kind of flexible collaboration, the possibility to leave a note for when others had time. The analyses further revealed how, in QueryTogether, most collaboration strategies typical of both conditions, such as, for example, the labor division and the establishment of common ground, could be more easily supported by the availability of the entities as searchable objects.

Common ground. Previous research demonstrated the importance of providing common ground in co-located interaction (Jacucci et al., 2009; Morrison et al., 2009). The higher lookup rate at the shared display suggests that better group awareness might have been achieved. Group awareness usually has a positive effect on the coordination of actions, anticipation, and assistance provided to collaborators with their local tasks (Gutwin & Greenberg, 2002). In particular, the higher awareness of publicly shared information suggests that the system facilitated the creation of a common understanding (Gutwin & Greenberg, 2002). This intuition is also supported by our results on the verbal usage of entities in QueryTogether, revealing how, in QueryTogether, more time was spent in creating common ground by sharing prior knowledge on entities, asking clarification questions, explaining, and reporting activities performed with entities. However, this sometimes led to long discussions that distracted from the main task. This suggests that the proper facilitation of common ground establishment is a challenge that requires further research.

Other implications of entity-based search. While this work suggests that entity-based search can effectively support collaborative information exploration, we believe that our findings may have implications for the design of search interfaces supporting a variety of search tasks beyond exploration. For example, even less explorative tasks include cases in which users search for a known item but do not recall either the title or any precise element from the content. In such cases, entity-based search may effectively support re-finding of such items by remembering an adjacent entity (e.g., an author, a similar document, or an imperfectly formulated topic), as making a query with the adjacent entity could trigger the system to recommend the desired item.

6.1. Open challenges

The work done in designing and studying QueryTogether uncovered some challenges that should be addressed in future research, which we describe below.

- *Supporting more direct access to entities.* Our design took a first step towards providing users with searchable objects related to their conversations. However, many entities related to the conversations were still missing from the proposed recommendations. The challenge for future research is understanding how entities can be extracted directly from the speech content and converted to searchable objects. This would require, among other things, the design of smart ways to extract real-time context-aware information about the conversation (McGregor & Tang, 2017), so that acceptable speech recognition accuracy could be achieved.
- *Identifying useful entity types.* The current implementation of QueryTogether has arbitrarily fixed entity types (i.e. documents, authors, and keywords). However, the most appropriate set of entity types depends on the specific scenario supported. In this case,

the Author type was not considered to be of much use to the participants (see Fig. 8), while some expressed interest in other possible types like the publication venue, for example. The challenge is understanding which entities to support in different situations.

- *Supporting more naturalistic tasks.* Our results reveal key insights and opportunities for how entity-centric exploration can be leveraged to support collaborative search with heterogeneous devices. While the proposed design proved to be effective on the task of collaborative literature search, understanding how to support more naturalistic tasks and informal conversations is a challenge that would require additional research.
- *Providing result previews on small devices.*Designing an interface for small devices involves making choices on what information to display, as the limited real estate of the devices' screen doesn't allow to show the same amount of information that is typically shown on interfaces designed for desktop environments. In QueryTogether, the preview of search results of scientific literature only included titles and authors of documents, leaving out information such as for example publication venue. While the same information was available in QueryTogether and the baseline, the difference in preview made a difference in how easily people could visually filter search results based on specific parameters. The challenge is to devise novel visualization strategies that would be able to fit the most important information with the least space possible, while maintaining legibility and skimmability.

6.2. Limitations

Evaluating collaborative search systems is a challenging task (Soulier, Tamine, Sakai, Azzopardi, & Pickens, 2016). Our study was a controlled within-groups user study in which the groups employed two system variants. This experimental design allowed for a direct comparison of the contribution of the various system variants and controlled for individual differences and behavioral patterns, which have been found to play a role in the perceived usefulness of search and recommender systems (Ekstrand, Harper, Willemsen, & Konstan, 2014; Hu & Pu, 2010). The advantages of a controlled study, however, come with limitations. The groups performed simulated work tasks with given topics, which may have affected the naturalness of user behavior. On the other hand, this mitigated the confounding factors potentially arising from the pre-knowledge of the groups about some given tasks.

The study used a single dataset from a single domain, which may limit the generalizability of our findings for scenarios involving data from other domains. On the other hand, the dataset was real, consisting of complete data from scientific articles consisting of over 50 million articles from several sources. This ensured that the coverage of the data was appropriate to support diverse search behavior and to allow for studying task performance with simulated work tasks.

We found that the QueryTogether system generally improves collaborative search in terms of several measures. However, we acknowledge that the contribution of the various features of the design were not investigated as independent factors in the present study. Consequently, our results suggest that multi-device co-located collaborative search can benefit from some or all of the features of the QueryTogether system. Future research should be aimed at confirming the effects of (1) input (entities vs. typed queries); (2) the collaboration mode (shared entity space vs. shared document space); and (3) devices (conventional typing devices vs. touch devices).

Finally, the questionnaires and qualitative analysis revealed participants' behavior, communication strategies, and collaboration styles that were not captured via the quantitative data-analysis. These analyses were based on qualitative and subjective feedback, and more research is required to confirm these results. However, despite these limitations, we believe that our quantitative and qualitative evidence on effectiveness, user engagement, contribution balance, and collaboration and communication styles show promise and open an interesting frontier for research targeting the improvement of the increasingly important field of collaborative information access.

7. Conclusions

Search is a social activity that pervades our daily life. The far reaching vision of this research is an environment able to understand our conversation, identifying key objects being referred to, and turning them into searchable entities. In this vision, search becomes a task carried out mostly by a system in the background, while users can focus their energy on discovery, sense making and collaboration. The assumption is that search is not the ultimate goal, but a mean for users who are engaged in solving complex tasks that transcend finding the information (Jacucci, 2016). With this work we took a promising first step in this direction. We introduced QueryTogether, a tool that facilitates collaborative search in spontaneous settings, by leveraging an entity-centric design based on entities as searchable objects and smart entity recommendations. In QueryTogether, most entities mentioned as part of a conversation were readily available as search objects that could directly be used in search without resorting to typing. Our system's novel features allowed us to obtain an environment effectively supporting the spontaneity of ubiquitous scenarios typical of our vision with the more effective usage of heterogeneous devices, a more flexible change of working style, and better search engagement. Moreover, QueryTogether supported effective collaboration through more balanced participation and by creating opportunities for common ground establishment. While there are still many open challenges regarding support for multi-device search in ubiquitous settings, our work highlights the benefit of interfaces based on entity-centric exploration and uncovers design opportunities for future developments of the approach.

Acknowledgments

This work has been partially supported by TEKES (Re:Know2). The funding source had no involvement in the study design, in the collection, analysis and interpretation of the data.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ipm.2018.04.005.

References

Amershi, S., & Morris, M. R. (2008). *Cosearch: a system for co-located collaborative web search. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1647–1656.

Andolina, S., Klouche, K., Peltonen, J., Hoque, M., Ruotsalo, T., Cabral, D., et al. (2015). *Intentstreams: smart parallel search streams for branching exploratory search. Proceedings of the 20th international conference on intelligent user interfaces*. ACM 300–305.

Andolina, S., Orso, V., Schneider, H., Klouche, K., Ruotsalo, T., Gamberini, L., et al. (2018). *Investigating proactive search support in conversation. Proceedings of the 2018 Conference on Designing Interactive Systems*. ACM http://dx.doi.org/10.1145/3196709.3196734.

Balog, K., Meij, E., & de Rijke, M. (2010). *Entity search: Building bridges between two worlds. Proceedings of the 3rd international semantic search workshop (SEMSEARCH '10)*. New York, NY, USA: ACM 9:1–9:5. http://dx.doi.org/10.1145/1863879.1863888.

Brown, B., McGregor, M., & McMillan, D. (2015). *Searchable objects: Search in everyday conversation. Proceedings of the 18th ACM conference on computer supported cooperative work & social computing (CSCW '15)*. New York, NY, USA: ACM 508–517. http://dx.doi.org/10.1145/2675133.2675206.

Capra, R., Chen, A. T., Hawthorne, K., Arguello, J., Shaw, L., & Marchionini, G. (2012). Design and evaluation of a system to support collaborative search. *Proceedings of the Association for Information Science and Technology, 49*(1), 1–10.

Castano, S., Ferrara, A., & Montanelli, S. (2014). *inwalk: Interactive and thematic walks inside the web of data. Proceedings of the Edbt*. Citeseer 628–631.

Chung, H., North, C., Self, J. Z., Chu, S., & Quek, F. (2014). Visporter: Facilitating information sharing for collaborative sensemaking on multiple displays. *Personal and Ubiquitous Computing, 18*(5), 1169–1186.

Clark, H. H., Brennan, S. E., et al. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition, 13*(1991), 127–149.

Dewan, P., & Choudhard, R. (1991). *Flexible user interface coupling in a collaborative system. Proceedings of the SIGCHI conference on human factors in computing systems (CHI '91)*. New York, NY, USA: ACM 41–48. http://dx.doi.org/10.1145/108844.108851.

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., et al. (2014). *Knowledge vault: A web-scale approach to probabilistic knowledge fusion. Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '14)*. New York, NY, USA: ACM 601–610. http://dx.doi.org/10.1145/2623330.2623623.

Ekstrand, M. D., Harper, F. M., Willemsen, M. C., & Konstan, J. A. (2014). *User perception of differences in recommender algorithms. Proceedings of the 8th ACM conference on recommender systems*. ACM 161–168.

Evans, B. M., & Chi, E. H. (2010). An elaborated model of social search. *Information Processing & Management, 46*(6), 656–678.

Firebaugh, G. (1999). Empirics of world income inequality. *American Journal of Sociology, 104*(6), pp. 1597–1630.

Golovchinsky, G., Adcock, J., Pickens, J., Qvarfordt, P., & Back, M. (2008). *[n. d.]. Cerchiamo: a collaborative exploratory search tool. Cscw*.

Golovchinsky, G., Dunnigan, A., & Diriye, A. (2012). *Designing a tool for exploratory information seeking. Chi'12 extended abstracts on human factors in computing systems*. ACM 1799–1804.

González-Ibáñez, R., Haseki, M., & Shah, C. (2013). 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. http://dx.doi.org/10.1016/j.ipm.2012.12.008.

Gutwin, C., & Greenberg, S. (2002). A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW), 11*(3), 411–446.

Halvey, M., Vallet, D., Hannah, D., Feng, Y., & Jose, J. M. (2010). An asynchronous collaborative search system for online video search. *Information Processing & Management, 46*(6), 733–748. http://dx.doi.org/10.1016/j.ipm.2009.11.007 Collaborative Information Seeking.

Han, S., Yue, Z., & He, D. (2015). Understanding and supporting cross-device web search for exploratory tasks with mobile touch interactions. *ACM Transactions on Information Systems, 33*(4), 16:1–16:34. http://dx.doi.org/10.1145/2738036.

Hartmann, B., Morris, M. R., Benko, H., & Wilson, A. D. (2009). *Augmenting interactive tables with mice & keyboards. Proceedings of the 22nd annual acm symposium on user interface software and technology*. ACM 149–152.

Hasida, K. (2007). Semantic authoring and semantic computing. In A. Sakurai, K. Hasida, & K. Nitta (Eds.). [*New Frontiers in Artificial Intelligence: JSAI 2003 and JSAI 2004 Conferences and Workshops*]. Berlin, Heidelberg: Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-71009-7_12 Niigata, Japan, June 23–27, 2003 and Kanazawa, Japan, May 31, - June 4, 2004, Revised Selected Papers.

Hearst, M. A. (2014). What's missing from collaborative search?Computer, (3), 58–61.

Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management, 44*(2), 957–962.

Hinrichs, U., Hancock, M., Carpendale, S., & Collins, C. (2007). *Examination of text-entry methods for tabletop displays. Proceedings of the second annual IEEE international workshop on horizontal interactive human-computer systems, 2007. TABLETOP '07*. IEEE 105–112.

Hu, R., & Pu, P. (2010). *A study on user perception of personality-based recommender systems. User modeling, adaptation, and personalization*. Springer 291–302.

Isaacs, E. A., Tang, J. C., & Morris, T. (1996). *Piazza: a desktop environment supporting impromptu and planned interactions. Proceedings of the 1996 ACM conference on computer supported cooperative work*. ACM 315–324.

Jacucci, G. (2016). Beyond search queries. *Computer, 49*(10), 90–93.

Jacucci, G., Peltonen, P., Morrison, A., Salovaara, A., Kurvinen, E., & Oulasvirta, A. (2009). *Ubiquitous media for collocated interaction. Shared encounters*. Springer 23–45.

Jain, A., Lupfer, N., Qu, Y., Linder, R., Kerne, A., & Smith, S. M. (2015). *Evaluating tweetbubble with ideation metrics of exploratory browsing. Proceedings of the 2015 ACM SIGCHI conference on creativity and cognition (C&C '15)*. New York, NY, USA: ACM53–62. http://dx.doi.org/10.1145/2757226.2757239.

Jeh, G., & Widom, J. (2003). *Scaling personalized web search. Proceedings of the. www (WWW '03)*. ACM 271–279. http://dx.doi.org/10.1145/775152.775191.

Jetter, H.-C., Gerken, J., Zöllner, M., Reiterer, H., & Milic-Frayling, N. (2011). *Materializing the query with facet-streams: A hybrid surface for collaborative search on tabletops. Proceedings of the chi '11*. New York, NY, USA: ACM3013–3022. http://dx.doi.org/10.1145/1978942.1979390.

Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., & Jacucci, G. (2015). *Designing for exploratory search on touch devices. Proceedings of the 33rd annual ACM conference on human factors in computing systems*. ACM 4189–4198.

Kraut, R. E., & Streeter, L. A. (1995). Coordination in software development. *Communications of the ACM, 38*(3), 69–82.

Krishnappa, R. (2005). *Multi-user search engine (muse): Supporting collaborative information seeking and retrieval* Master Thesis.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM, 49*(4), 41–46. http://dx.doi.org/10.1145/1121949.1121979.

Marie, N., Gandon, F., Ribière, M., & Rodio, F. (2013). *Discovery hub: on-the-fly linked data exploratory search. Proceedings of the 9th international conference on semantic systems*. ACM 17–24.

McGregor, M., & Tang, J. C. (2017). *More to meetings: Challenges in using speech-based technology to support meetings. Proceedings of the 2017 acm conference on computer supported cooperative work and social computing (CSCW '17)*. New York, NY, USA: ACM 2208–2220. http://dx.doi.org/10.1145/2998181.2998335.

Miller, J., Sumeeth, M., & Singh, R. I. (2011). Evaluating the readability of privacy policies in mobile environments. *International Journal of Mobile Human Computer Interaction, 3*(1), 55–78. http://dx.doi.org/10.4018/jmhci.2011010104.

Morris, M. R. (2008). *A survey of collaborative web search practices. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1657–1660.

Morris, M. R. (2013). *Collaborative search revisited. Proceedings of the 2013 conference on computer supported cooperative work*. ACM 1181–1192.

Morris, M. R., Fisher, D., & Wigdor, D. (Fisher, Wigdor, 2010a). Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks.

Information Processing & Management, 46(6), 703–717.

Morris, M. R., & Horvitz, E. (2007). *Searchtogether: an interface for collaborative web search. Proceedings of the 20th annual ACM symposium on user interface software and technology*. ACM 3–12.

Morris, M. R., Lombardo, J., & Wigdor, D. (Lombardo, Wigdor, 2010b). *Wesearch: supporting collaborative search and sensemaking on a tabletop display. Proceedings of the 2010 acm conference on computer supported cooperative work*. ACM 401–410.

Morris, M. R., Paepcke, A., & Winograd, T. (2006). *Teamsearch: comparing techniques for co-present collaborative search of digital media. Proceedings of the tabletop '06*. Washington, DC, USA: IEEE Computer Society 97–104. http://dx.doi.org/10.1109/TABLETOP.2006.32.

Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., et al. (2009). *Like bees around the hive: a comparative study of a mobile augmented reality map. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1889–1898.

Nuzzolese, A. G., Presutti, V., Gangemi, A., Peroni, S., & Ciancarini, P. (2017). Aemoo: Linked data exploration based on knowledge patterns. *Semantic Web, 8*(1), 87–112.

O'Brien, H. L., & Toms, E. G. (2013). Examining the generalizability of the user engagement scale (ues) in exploratory search. *Information Processing & Management, 49*(5), 1092–1107 https://doi.org/10.1016/j.ipm.2012.08.005.

Paul, S. A., & Morris, M. R. (2009). *Cosense: enhancing sensemaking for collaborative web search. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1771–1780.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008). *Algorithmic mediation for collaborative exploratory search. Proceedings of the SIGIR '08*. New York, NY, USA: ACM 315–322. http://dx.doi.org/10.1145/1390334.1390389.

Rogers, Y., kyung Lim, Y., PhD, W. R. H., & Marshall, P. (2009). Equal opportunities: Do shareable interfaces promote more group participation than single user displays? *HumanComputer Interaction, 24*(1-2), 79–116. http://dx.doi.org/10.1080/07370020902739379.

Ruotsalo, T., Jacucci, G., Myllymäki, P., & Kaski, S. (2014). Interactive intent modeling: Information discovery beyond search. *Communications of the ACM, 58*(1), 86–92.

Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*(1), 89–99.

Scott, S. D., Carpendale, M. S. T., & Inkpen, K. M. (2004). *Territoriality in collaborative tabletop workspaces. Proceedings of the 2004 ACM conference on computer supported cooperative work*. ACM 294–303.

Shah, C. (2010). Coagmento - a collaborative information seeking, synthesis and sense-making framework. *Integrated demo at CSCW, 2010*.

Shiga, S., Joho, H., Blanco, R., Trippas, J., & Sanderson, M. (2017). *Modelling information needs in collaborative search conversations. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. New York, NY, USA: ACM 715–724. http://dx.doi.org/10.1145/3077136.3080787.

Soulier, L., Tamine, L., Sakai, T., Azzopardi, L., & Pickens, J. (2016). Report on the first international workshop on the evaluation on collaborative information seeking and retrieval (ecol'2015). *SIGIR Forum, 50*(1), 42–48. http://dx.doi.org/10.1145/2964797.2964805.

Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: Computer support for collaboration and problem solving in meetings. *Communications of the ACM, 30*(1), 32–47.

Tang, A., Tory, M., Po, B., Neumann, P., & Carpendale, S. (2006). *Collaborative coupling over tabletop displays. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1181–1190.

Teevan, J., Morris, M., & Azenkot, S. (2014). Supporting interpersonal interaction during collaborative mobile search. *Computer, 47*(3), 54–57. http://dx.doi.org/10.1109/MC.2013.449.

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin, 63*(6), 384.

Tunkelang, D. (2009). Faceted search. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 1*(1), 1–80.

White, R. W., & Roth, R. A. (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 1*(1), 1–98.

Williams, K. D., & Karau, S. J. (1991). Social loafing and social compensation: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology, 61*(4), 570.

Wilson, M. L., Kules, B., Schraefel, m.c., & Shneiderman, B. (2010). *From keyword search to exploration: Designing future search interfaces for the web2. From keyword search to exploration: Designing future search interfaces for the web* Found Trends Web Science 1–97. http://dx.doi.org/10.1561/1800000003.

Wiltse, H., & Nichols, J. (2009). *Playbyplay: collaborative web browsing for desktop and mobile devices. Proceedings of the SIGCHI conference on human factors in computing systems*. ACM 1781–1790.

Yee, K.-P., Swearingen, K., Li, K., & Hearst, M. (2003). *Faceted metadata for image search and browsing. Proceedings of the SIGCHI conference on human factors in computing systems (CHI '03)*. New York, NY, USA: ACM 401–408. http://dx.doi.org/10.1145/642611.642681.

Yue, Z., Han, S., He, D., & Jiang, J. (2014). Influences on query reformulation in collaborative web search. *Computer, 47*(3), 46–53.

Zhai, C., & Lafferty, J. (2004). A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems, 22*(2), 179–214. http://dx.doi.org/10.1145/984321.984322.

Zhai, C. X., Cohen, W. W., & Lafferty, J. (2003). *Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. Proceedings of the 26th annual international ACM SIGIR conference on research and development in informaion retrieval (SIGIR '03)*. New York, NY, USA: ACM 10–17. http://dx.doi.org/10.1145/860435.860440.

Zhang, Y., Pfeuffer, K., Chong, M. K., Alexander, J., Bulling, A., & Gellersen, H. (2017). Look together: Using gaze for assisting co-located collaborative search. *Personal and Ubiquitous Computing, 21*(1), 173–186. http://dx.doi.org/10.1007/s00779-016-0969-x.

Publication V

Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, (C&C '15). ACM, New York, NY, USA, 103-106, 2015. DOI: 10.1145/2757226.2757252, June 2015.

InspirationWall: Supporting Idea Generation Through Automatic Information Exploration

Salvatore Andolina[1], **Khalil Klouche**[1,3], **Diogo Cabral**[1], **Tuukka Ruotsalo**[2], **and Giulio Jacucci**[1,2]
[1]Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki, Finland
[2]Helsinki Institute for Information Technology HIIT, Aalto University, Finland
[3]Helsinki Media Lab, Aalto University, Finland

ABSTRACT
Collaborative idea generation leverages social interactions and knowledge sharing to spark diverse associations and produce creative ideas. Information exploration systems expand the current context by suggesting novel but related concepts. In this paper we introduce InspirationWall, an unobtrusive display that leverages speech recognition and information exploration to enhance an ongoing idea generation session with automatically retrieved concepts that relate to the conversation. We evaluated the system in six idea generation sessions of 20 minutes with small groups of two people. Preliminary results suggest that InspirationWall contrasts the decay of idea productivity over time and can thus represent an effective way to enhance idea generation activities.

Author Keywords
Idea generation; Information Exploration; Automatic Speech Recognition.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

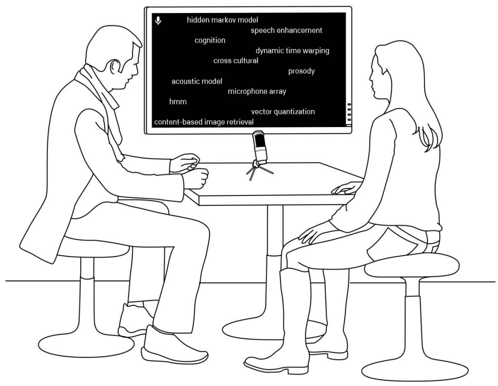


Figure 1. InspirationWall is an unobtrusive display supporting idea generation by leveraging speech recognition and automatic information exploration. It monitors users' discussion and automatically suggests keywords to support their idea generation.

INTRODUCTION
Collecting and navigating through information is an important phase in creative processes [13], which fosters associative and inspirational learning [2]. Previous work that sought to support for example brainstorming referred to the semantic network structure of human memory, where concepts feature as nodes with associative links [15]. In brainstorming, one cognitive operation to generate ideas is to retrieve concepts from associative memory. Expanding the current context of topics has been investigated through topic suggestion algorithms designed to generate candidate topics that are novel but related to the current context [9]. As brainstorming is often a collaborative practice, recent creativity systems support groups. Groups generally perform better than individuals in a variety of tasks [6]. Group brainstorming can be effective in generating creative ideas as suggested by cognitive approaches [3], and technology may help minimizing the effect of negative social processes [5]. A beneficial feature in group brainstorming is the ability to detect the context and content of the brainstorming through utterances of participants. IdeaExpander [14] is a tool to support group brainstorming by intelligently selecting pictorial stimuli based on the group's conversation on a chat. The pictures generally enhanced performance as measured by both originality and diversity of ideas [15]. Less investigated are face to face systems in group sessions that suggest keywords instead of pictures. Systems suggesting keywords and topics have recently been applied successfully to improve exploratory search processes [1, 10]. Such systems predict the current intent model of the user in the exploration and suggest possible explorations. These approaches have also been found useful in avoiding keyword input by selecting and manipulating suggested keywords by touch [8]. The present work investigate further alternative input modalities such as speech to text that permit the system to run in the background without interrupting the creative process but providing a continuous resource.

C&C '15, June 22–25, 2015, Glasgow, United Kingdom
 ISBN 978-1-4503-3598-0/15/06...$15.00.
DOI: http://dx.doi.org/10.1145/2757226.2757252

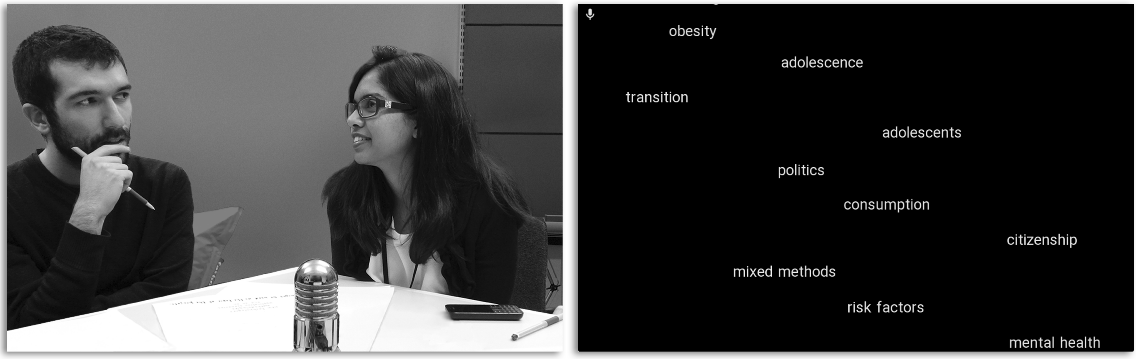


Figure 2. Left: Two participants in a brainstorming session. Right: A screen capture from the InspirationWall interface.

SYSTEM DESIGN

InspirationWall is a non-intrusive source of diverse ideas (Figure 1). It was designed as a low-key visual aid, as to not interfere with the user's idea generation process. InspirationWall continuously monitors the discussions through a conference microphone and the input to the system is recorded from users natural interaction via speech recognition. Speech recognition is performed in real-time using Googles implementation of the HTML5 Web Speech API [16].

Recognized expressions are processed by an entity-based keyword suggestion system that returns related keywords by discovering associated and novel information related to the input [11]. Returned keywords are then displayed as slowly crossing the screen from top to bottom as to allow a progressive refreshing of displayed keywords. The graphical interface of Inspiration Wall is minimal: it runs fullscreen with a black background. Keywords are displayed in white. Every two seconds providing the keyword buffer is not empty a new keyword appears at the top of the screen at a random horizontal position and falls slowly towards the bottom of the screen.

Keyword Suggestion System

As the set of potential keywords matching any part of users' discussion is likely to be much higher than what can be presented for the user, and the discussion can contain misleading cues due to the natural dialogues that the system listens, we use a centrality-based ranking of the keywords in a large knowledge-graph.

Intuitively, this approach allows the system to suggest central keywords that are related to the user input via the knowledge-graph rather than only suggesting keywords directly matching to the input. This can help discovering keywords that are highly relevant for the input, but at the same time central to the overall discussion [12]. The knowledge-graph G is undirected and labeled and consists of a disjoint union set of keywords and documents (called nodes $n \in G$) and the set of edges between the documents and the keywords. Each keyword in the graph is connected to a document it describes. For example, an article about "relevance feedback for information retrieval" could be described with a set of keywords, such as "information retrieval", "relevance feedback", "implicit feedback", "web search", and so on. In addition, we index the text of the articles that is used to retrieve an initially relevant set of documents from which the knowledge-graph is constructed.

The user's query consists of one or more words detected via the speech recognition system. A set of keywords detected are called preference keywords $q \in G$ in the graph, where $|q| = 1$ and q_j denotes the preference for keyword j. In our case no weighting is conducted for the keywords so the preference is uniformly distributed over for the given keywords in q.

We use the personalized PageRank method [7] to compute the ranking of the nodes given the q. It can be then formalized as follows. Let an individual node be denoted as n, and by $I(n)$ and $O(n)$ denote the set of in-neighbors and out-neighbors of n in G respectively. Let A be the matrix corresponding to the graph G, where

$$A_{ij} = \frac{1}{|O_{ij} \cup I_{ij}|}$$

if the node i links to the node j or vice versa, and $A_{ij} = 0$ otherwise. For a given q, the personalized PageRank equation can be written as

$$v = (1 - c)Av + cq,$$

where $c = 0.15$ is the teleportation rate. The solution v is a steady-state distribution of random surfers, where a surfer teleports at each step to a node n with probability $c \cdot q(n)$, or moves to a random neighbor otherwise. We compute the steady distribution by using the power iteration method with 100 iterations.

The weights of the v are directly used in ranking the keywords. As the size of our knowledge-graph is hundreds of millions of nodes, the computation is not possible on-line for the complete graph. To make the PageRank computation feasible with an acceptable latency, we approximate the set of nodes to be included in the initial graph by using a language

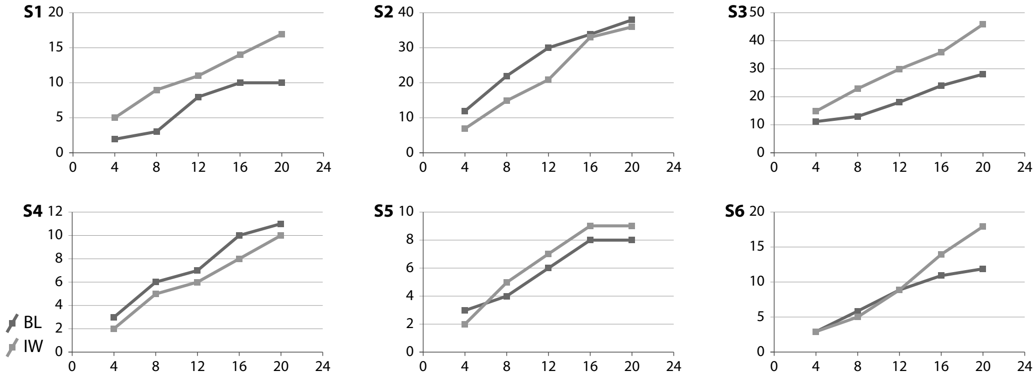


Figure 3. Accumulation of ideas per condition (BL = Baseline; IW = InspirationWall) in the different sessions S1,...,S6. On the Y-axis is the cumulative number of ideas, and on the X-axis is the time from the beginning of the session (minutes).

model approach of information retrieval [17] and select 3000 documents and the corresponding entities to be cumulatively added in the knowledge-graph at each iteration.

EVALUATION

We designed an experiment to evaluate the effect of the system on the idea generation process. The goal of the experiment was (1.) to understand if and how InspirationWall helped small groups generating more ideas, and (2.) to assess the overall effectiveness of the system as a creativity support tool using standard metrics.

Participants

The evaluation was conducted in groups of two persons (Figure 2). We recruited twelve participants (six pairs) with experience in idea generation activities from the computer science departments of two universities. Three of the participants were females and the mean age was $M = 28.33$, $SD = 3.98$. To simulate more natural discussions and brainstorming activity, we ensured that participants in the same pair knew each other. Participants were non-native English speakers from different countries and cultures (Iran, Canada, Spain, Nepal, Italy, Turkey, Sri Lanka, Rwanda, Kenya) with a similar level of proficiency in oral English. Their levels of education were: 25% PhD, 67% Master, 8% Bachelor. Participants received two movie tickets as a compensation for their participation.

Tasks

We used a within-group design, where groups were asked to perform two tasks: one with the support of InspirationWall and one without external support. We counterbalanced by changing the order in which the two tasks were performed and the order in which the groups were subjected to each condition.

The task was created to support an idea generation scenario and formulated as follows: *Imagine you have to come up with novel student projects on topic X. Please generate as many ideas as possible for new technologies, interaction techniques, methodologies, application scenarios, and so on, that might be used as more specific topics of the projects on topic X.* Two topics were used in the evaluation sessions: (1.) Robotics, and (2.) Wearable computing.

Metrics and Results

Quantity of Ideas

Since we were interested to check whether our application influenced the number of ideas generated, we have looked to the cumulative number of ideas considering time and session (Table 1). In total, the six groups have produced 107 ($M = 30.57$, $SD = 12.24$) ideas without external support and 136 ($M = 38.86$, $SD = 14.99$) using InspirationWall. In Figure 3, it is shown the accumulation of ideas per condition in the sessions, considering intervals of 4 minutes. In addition, video recordings obtained from the camera placed between the participants and the InspirationWall display allowed us to count the occurrences of participants looking at the screen (results shown in Figure 4). It is interesting to observe that the three groups (S1, S3 and S6) that have looked at the display the most, improved their performance with respect to the baseline condition, presenting a higher number of generated ideas and a more constant productivity.

Creativity Support Index

To measure the performance of our system in terms of creativity support, we involved participants in the assessment of the Creativity Support Index (CSI) [4].

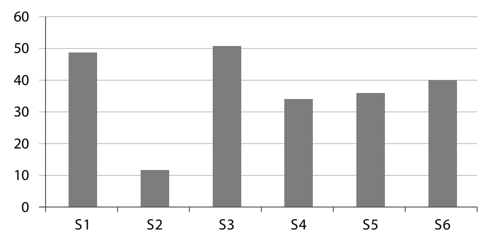


Figure 4. Number of occurrences of participants looking at the screen in sessions S1,...,S6.

Table 1. **Accumulation of ideas per condition (BL = Baseline; IW = InspirationWall) at time $T = 4, 8, 12, 16, 20$ in sessions S1,...,S6. For each point in time, p-values from paired t-tests are also shown.**

T		S1	S2	S3	S4	S5	S6	M (SD)	p
4	BL	2	12	11	3	3	3	5.67 (4.55)	1
	IW	5	7	15	2	2	3	5.67 (4.97)	
8	BL	3	22	13	6	4	6	9.00 (7.27)	0.61
	IW	9	15	23	5	5	5	10.33 (7.34)	
12	BL	8	30	18	7	6	9	13.00 (9.38)	0.73
	IW	11	21	30	6	7	9	14.00 (9.51)	
16	BL	10	34	24	10	8	11	16.17 (10.48)	0.23
	IW	14	33	36	8	9	14	19.00 (12.30)	
20	BL	10	38	28	11	8	12	17.83 (12.24)	0.17
	IW	17	36	46	10	9	18	22.67 (15.00)	

This index is computed from two sets of six questions and each question related with a factor. The six factors that compose the CSI are: *Collaboration*, *Enjoyment*, *Exploration*, *Expression*, *Expressiveness*, *Immersion* and *Results Worth Effort*. Each pair of questions are weighted based in pair wise comparisons of the factors made by each participant. The result of the CSI was $M = 53.36$, $SD = 13.35$. The most important factors for the participants were *Expressiveness* ($M = 3.58$, $SD = 1.24$) and *Exploration* ($M = 3.83$, $SD = 0.94$).

DISCUSSION AND CONCLUSIONS

Creative ideas are often triggered by unexpected associations. InspirationWall offers a quiet additional source of information to fuel the activity of collaborative idea generation. This paper presents the implementation and a preliminary evaluation of such a system. Participants were asked to generate ideas but not explicitly to use or interact with the system which was simply provided as is. Our study shows that participants that used InspirationWall more – as indicated by the count and duration of gazing occurrences obtained through video analysis – tended to generate more ideas in total and over time. Those results suggest that InspirationWall contrasts the decay of idea productivity over time typical of traditional idea generation sessions. Although the CSI does not show a high value, it is still above the median value of the scale, with the most important factors for the participants being *Expressiveness* and *Exploration*. Such results confirms the effectiveness of automatic information exploration and keyword suggestion on idea generation, opening a variety of directions for future work, including for example application to other datasets, and allowing richer interactions with the system through touch. The novel approach on idea generation support described in this paper, the simple design of our prototype and the positive results of this preliminary study are the contributions of our work to the future of digital tools for creativity support.

ACKNOWLEDGEMENTS

The data used in the experiments is derived from the Web of Science prepared by THOMSON REUTERS, Inc., and the digital libraries of ACM, IEEE, and Springer. This work has been partially supported by TEKES (Re:Know) and the Academy of Finland (278090).

REFERENCES

1. Andolina, S., et al. Intentstreams: smart parallel search streams for branching exploratory search. In *Proc. IUI*, ACM (2015).
2. Binder, T., Ehn, P., De Michelis, G., Jacucci, G., Linde, G., and Wagner, I. *Design things*. MIT Press Cambridge, MA, 2011.
3. Brown, V. R., and Paulus, P. B. Making Group Brainstorming More Effective: Recommendations From an Associative Memory Perspective. *Current Directions in Psychological Science 11*, 6 (Dec. 2002), 208–212.
4. Cherry, E., and Latulipe, C. Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput.-Hum. Interact. 21*, 4 (June 2014), 21:1–21:25.
5. Diehl, M., and Stroebe, W. Productivity loss in brainstorming groups: Toward the solution of a riddle. *J Pers Soc Psycho* (1987), 497–509.
6. Hill, G. W. Group versus individual performance: Are n+1 heads better than one? *Psychological bulletin 91*, 3 (1982), 517–539.
7. Jeh, G., and Widom, J. Scaling personalized web search. In *Proc. WWW*, WWW '03, ACM (2003), 271–279.
8. Klouche, K., Ruotsalo, T., Cabral, D., Andolina, S., Bellucci, A., and Jacucci, G. Designing for exploratory search on touch devices. In *Proc. CHI*, ACM (2015).
9. Maguitman, A., Leake, D., and Reichherzer, T. Suggesting novel but related topics: towards context-based support for knowledge model extension. In *Proc. IUI*, ACM (2005), 207–214.
10. Ruotsalo, T., et al. Directing exploratory search with interactive intent modeling. In *Proc. CIKM*, ACM (2013), 1759–1764.
11. Ruotsalo, T., et al. Interactive intent modeling: Information discovery beyond search. *Commun. ACM 58*, 1 (Jan. 2015), 86–92.
12. Ruotsalo, T., and Hyvönen, E. Exploiting semantic annotations for domain-specific entity search. In *Proc. ECIR*, Springer (2015).
13. Shneiderman, B. Creativity support tools. *Commun. ACM 45*, 10 (Oct. 2002), 116–120.
14. Wang, H.-C., Cosley, D., and Fussell, S. R. Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli. In *Proc. CSCW*, ACM (2010), 103–106.
15. Wang, H.-C., Fussell, S. R., and Cosley, D. From diversity to creativity: Stimulating group brainstorming with cultural differences and conversationally-retrieved pictures. In *Proc. CSCW*, ACM (2011), 265–274.
16. Web Speech API. `https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html`.
17. Zhai, C., and Lafferty, J. A study of smoothing methods for language models applied to ad hoc information retrieval. In *Proc. SIGIR*, ACM (2001), 334–342.

Publication VI

Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Conference on Designing Interactive Systems*, (DIS '18). ACM, New York, NY, USA, 1141-1152, 2018. DOI: 10.1145/3064663.3064728, June 2018.

Investigating Proactive Search Support in Conversations

Salvatore Andolina[1], Valeria Orso[2], Hendrik Schneider[3,4], Khalil Klouche[3], Tuukka Ruotsalo[3], Luciano Gamberini[2], Giulio Jacucci[1,3]

[1]Helsinki Institute for Information Technology HIIT,
Department of Computer Science, Aalto University, Finland
[2]Human Inspired Technology Research Centre HIT, University of Padova, Italy
[3]Helsinki Institute for Information Technology HIIT,
Department of Computer Science, University of Helsinki, Finland
[4]University of Göttingen, Göttingen, Germany

ABSTRACT
Conversations among people involve solving disputes, building common ground, and reinforce mutual beliefs and assumptions. Conversations often require external information that can support these human activities. In this paper, we study how a spoken conversation can be supported by a proactive search agent that listens to the conversation, detects entities mentioned in the conversation, and proactively retrieves and presents information related to the conversation. A total of 24 participants (12 pairs) were involved in informal conversations, using either the proactive search agent or a control condition that did not support conversational analysis or proactive information retrieval. Data comprising transcripts, interaction logs, questionnaires, and interviews indicated that the proactive search agent effectively augmented the conversations, affected the conversations' topical structure, and reduced the need for explicit search activity. The findings also revealed key challenges in the design of proactive search systems that assist people in natural conversations.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords
Spoken conversation support; proactive search; voice interfaces; background speech.

INTRODUCTION
Casual collaborative tasks, such as travel planning or selecting a movie to watch create decision-making processes to solve disputes, build common ground, and reinforce mutual beliefs and assumptions based on people's preferences and prior knowledge. This process is conducted in a conversational

DIS '18, June 9–13, 2018, Hong Kong

ACM ISBN 978-1-4503-5198-0/18/06...$15.00
DOI: https://doi.org/10.1145/3196709.3196734

Figure 1. An illustration of the proactive search from spoken conversations. 1) The system listens to a natural spoken conversation between the participants; 2) information needs arise during the conversation; 3) the system is able to proactively retrieve useful information to support the conversation; 4) information needs are satisfied seamlessly and the conversation can continue smoothly.

exchange between two or more people, and a spoken form is often used to enable such an exchange. The key technologies assisting individuals in the decision-making processes are information retrieval and recommender systems, as well as question answering and summarization techniques [2, 34]. These systems can provide people with additional information to support their decisions and guide them to information that is important in the decision-making process.

Today, the availability of a wide array of personal devices has made search and recommendation possible in a variety of casual situations, including everyday conversations [9]. The ability to quickly conduct searches when co-located with others–a collaborative search–is believed to account for more than 60% of mobile searches [9, 45]. Often, however, search systems are not tailored to support the conversational process as such, but require explicit commands, preferences input, queries, and human attentional resources to guide the process [33]. As a consequence, the systems may disrupt social interactions rather than supporting them [1, 30, 32].

Proactive searches can leverage information from peoples' contexts to retrieve information in an easily accessible and non-intrusive manner [36]. Despite the current limitations of automatic speech recognition, recent research on voice-based interaction [9, 27] has shown that relevant contextual information can be extracted even from a partial recognition.

These findings suggest that there are opportunities for scenarios in which a search is proactively performed in the background by using naturally occurring spoken conversational contexts between individuals while the individuals stay focused on their conversations and pay attention to their personal devices only when they need additional information. As opposed to previous approaches, which have mainly focused on supporting creativity [40], using voice or conversation recordings as research tools to uncover what users are doing with their mobile devices [27], or understanding information needs in conversations as parts of search tasks [41], we aim to go a step further by investigating how to proactively retrieve information from the Web to augment conversations (Figure 1).

In detail, *we aim to investigate how a proactive search agent can support natural spoken conversations between people by augmenting the conversations with additional information*. More specifically, we investigate the following research questions:

RQ1 Does a proactive search system with spoken input from a natural conversation influence the conversation?

RQ2 Does a proactive search system with spoken input from a natural conversation affect the consumption of Web resources during conversations?

RQ3 Does a proactive search system with spoken input from a natural conversation affect participants' subjective experiences?

To investigate our research questions, we designed SearchBot, an agent system that listens to a conversation, detects entities mentioned in the conversation, and proactively retrieves and presents information related to the conversation. We used SearchBot in a study with 24 participants (12 pairs) engaging in informal discussions on building travel or movie lists. Our findings suggest that SearchBot can effectively augment the conversations by enriching them with entities and documents shown on the screen and allowing people to consult the same number of Web resources as they would with traditional explicit searches. Subjective data from interviews and questionnaires suggest that proactive search support was generally found to be useful but also revealed key challenges for the design of proactive search systems assisting people in natural spoken conversation.

BACKGROUND

Information retrieval has traditionally relied on the query-response paradigm, with the user expressing information needs as explicit queries and the search engine responding with information items estimated to fulfill the user's information needs. Despite its practical success in Web search engines, the interactions between the user and the search engine may be laborious, as the broader search context, the user's exact search intent, and evolving information needs can be difficult to capture without explicit user involvement [48].

Proactive search refers to an information retrieval setting where the search system tries to automatically or proactively *anticipate* the user's upcoming queries and information needs [12, 16, 25]. An early attempt to build a proactive search system is the *Remembrance Agent* [35], which indexes a user's personal data such as emails and written notes. The system runs continuously in the background and displays a list of document summaries related to the current document being read or written. *Letizia* [24] is another early system that provides proactive recommendations during Web browsing using a set of heuristic rules. Another example of using search history is proposed in [42], where patterns repeated over time are extracted and used to proactively recommend resources to the user at specific times of the day. Recent work has also deployed computer vision to automatically detect broader user context and infer the user's work tasks for which supporting information is retrieved proactively [46, 47]. An interesting application for proactive search that comes close to our approach utilizes the subtitles of TV broadcasts being viewed by the user as context for predicting potential information needs [20].

Commercially deployed examples include *Google Now* and *Microsoft Cortana*, which run on users' smartphones and provide resources based on the current context. In particular, Google Now tries to model not only short-term search intents but long-term interests and habits based on several months of search log data [18].

A related research area is search personalization, where the search engine tries to discern individuals' unique search goals [43, 44]. Typical techniques include anticipating users' needs by taking into account their search histories [3], pre-search context information [23], social networks [11], and interaction behaviors [19, 37]. Researchers have also used previous search queries by the same user as context [10] and considered various kinds of relationships between the subsequent queries as features, such as reformulating the query or narrowing the search scope [49].

Supporting Conversations

Conversational systems have been studied from various angles, often referred to as situated interaction [8, 13]. Challenges of situated interactions include, for example, modeling initiatives in interaction, contextual interpretation, grounding, and turn-taking.

Related work has also investigated various kinds of visualizations to support conversations. *ConversationCluster* [7] uses visualizations to highlight salient moments in live conversation while archiving a meeting. Similarly, *Second Messenger* [14] uses a speech-recognition engine as an input method and shows filtered keywords from the group's conversation on an interactive display with the goal of increasing the visibility of diverse viewpoints. Other work [15] shows individual speaker-participation rates on a shared display to influence group behavior during a conversation.

Another main stream of research on conversation support systems has focused on creative design discussions. Schiavo at al. [39] introduced a system that monitors group members' non-verbal behaviors and promotes balanced participation, giving the participants targeted directives through peripheral displays. *InspirationWall* leverages speech recognition and information exploration to augment a creative conversation with keywords that relate to the speech stream [4]. Similarly, *IdeaWall* [40] provides visual stimuli to the participants of a brainstorming session to facilitate the creative process. *Crowdboard* augments brainstorming conversations with real-time creative input from online crowds [6].

In this work, we aim to support a wider range of informal everyday conversations by augmenting them with information retrieved proactively by a search agent that listens to conversations.

Using Background Speech for Interaction

Speech-based interaction has been thoroughly studied in the literature. However, the interest in speech-based systems seems to have risen again in recent years, probably due to the recent advances in automatic speech recognition [31]. In particular, a large body of work focuses on a dialogic mode of interaction [28] where users communicate with the system using natural language. Commercially available examples include *Apple's Siri*, *Microsoft's Cortana*, and *Google Now*.

Less investigated is the use of background speech for interaction. One example is *Ambient Spotlight* [21], which uses speech recognition during meetings to search for desktop documents and puts them in a folder associated with the calendar entry related to that meeting. Other systems use background speech to retrieve words and other kinds of visual stimuli to support a creative conversation [4, 40]. As opposed to those systems, which are designed to support creative conversations where even misrecognitions and random results may lead to useful stimuli [22], we investigate how to support more generic conversations by proactively retrieving richer sources of information, such as documents, from the Web.

An important study related to our work is that of McGregor and Tang [26]. The aim of their study was to understand how well a speech-based agent could detect useful actions during formal meetings. Although the study used a simulated system to create a best-case scenario, results showed that extracted action items failed to fit with the meeting or gave an incorrect summary of what was being discussed or what the participants intended. A different approach was that of McMillan et al. [27]. Their study suggested that a continuous speech stream, rather than containing directly actionable items, can be used to identify users' next actions such as searches. This result inspired our research, as it means that regardless of the limitation of current automatic speech recognition technology, many useful words that would likely be used for a search could still be recognized. In this study, we aim to understand whether performing those searches proactively during conversations could effectively enrich those conversations.

THE PROACTIVE SEARCH SYSTEM

We designed the SearchBot system to monitor a conversation and provide continuous recommendations of related documents and entities in a non-intrusive way. Below, we describe the system's main components: spoken conversation analysis, user interface design, and recommendation and retrieval methodology.

Spoken Conversation Analysis

SearchBot listens to conversations through a microphone. Speech recognition is performed by using Google's implementation of the HTML5 Web Speech API[1]. The speech API takes an audio recording as an input and outputs a transcript in natural language. The speech recognizer is continuously listening to the conversation. The voice activity is automatically detected based on the audio input, and the system starts building a sentence from the input. After the activity stops, the system returns the recognized sentence. As soon as the system recognizes and returns the sentence transcript, it triggers the entity detection and recommendation component.

Entity Detection and Recommendation

Each transcription is processed by Google's Cloud Natural Language API[2], which is used to extract recognized entities from the transcripts. The API returns entities along with the information about their named entity types. For example, people, locations, and organizations are separately typed.

In order to recommend new entities based on the detected entities, we model them using a vector space model [38]. To recommend new entities, we train an entity embedding model using Word2Vec [29] on a complete English Wikipedia. The detected entities from the present transcript are each represented as a vector in the embedding space. The embedding model is used by first combining the vectors of the words in the recognized entities and then retrieving new entities by ranking other entities using their cosine similarity in the embedding space. Altogether, four highest-ranking entities are retrieved in response to each transcript (Figure $2c_2$). For example, for the input “Bordeaux”, “France”, and “wines”, the system computes a cosine distance for an input vector “Bordeaux” + “France” + “wines” and retrieves the entities “Bandol,” “sauternes,” “wines,” and “Marseille,” which have the smallest cosine distance to that vector.

Document Retrieval

Related documents are retrieved via Google Custom Search by combining entities recognized in the present transcript to a query. Entities of type “location” or “person” are prioritized to improve the relevance of the shown results. If an entity of such type is identified, a separate query is generated using that named entity, and the other entities are combined to that query. Altogether, four search results are retrieved in response to each transcript (Figure $2c_1$).

More specifically, anytime the recognizer detects pauses, silence, or non-speech audio, a new sentence is returned. From the sentence, a set of entities is extracted, and a type is determined for each entity. If some of those entities are named entities, in our case of type “location” or “person,” they are stored in a separate named entity query vector. All the entities are also used to form another general query vector. The final

[1] https://w3c.github.io/speech-api/speechapi.html

[2] https://cloud.google.com/natural-language/

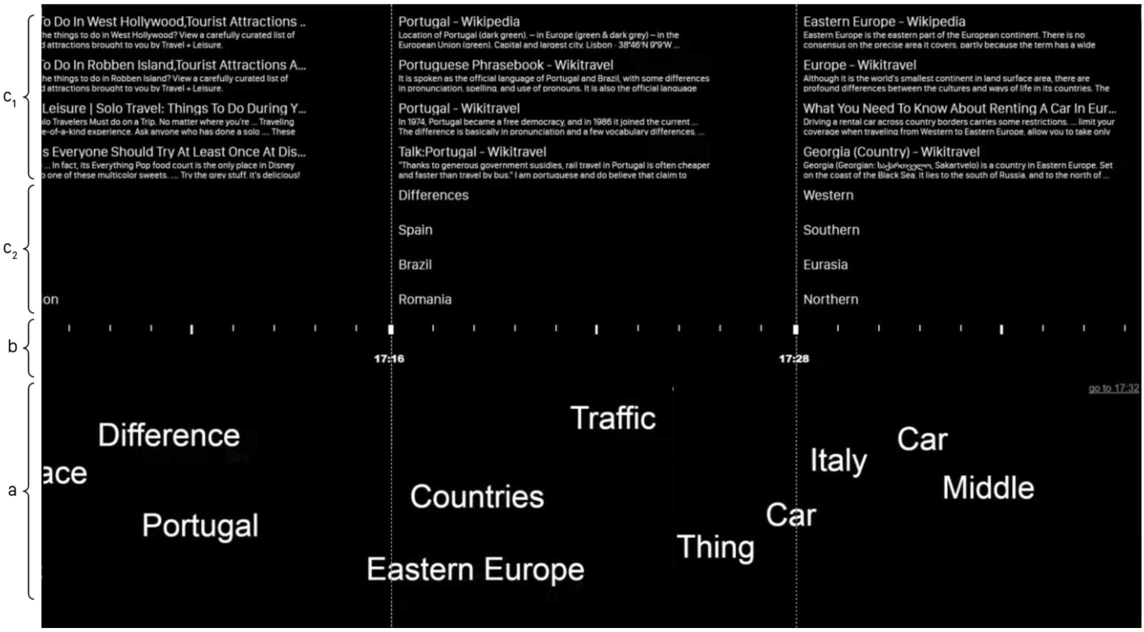


Figure 2. The user interface of the SearchBot system. The system monitors a conversation and provides continuous recommendations of related documents and entities in a non-intrusive way. a) Stream of recognized entities; b) timescale with timecodes; c_1) recommended documents; c_2) recommended entities.

set of search results shown to the user is then computed as the union of the highest ranked results in response to both query vectors. In case none of the entities are of type location or person, only the latter vector is used for retrieval.

An example of a sentence, extracted entities and their types, and the query vectors is given below:

SENTENCE: `Bordeaux is famous for its wines.`
ENTITIES: `Bordeaux (type` *location*`), wines (type` *consumer good*`)`
NAMED ENTITY QUERY VECTOR: `Bordeaux`
GENERAL QUERY VECTOR: `Bordeaux + wines`

User Interface Design

The user interface operated on a regular Web browser. It consists of a timeline that displays a stream of recognized entities in the lower part of the window (Figure 2a), a timescale with timecodes displayed in the center (Figure 2b), and successive sets of four retrieved documents (Figure 2c_1) and four recommended entities in the upper part of the window (Figure 2c_2). A new set extends the timeline every time a new transcription is available.

The user can interact with the system in multiple ways. Clicking on recognized or recommended entities triggers a search and opens the most relevant article in a new tab. Clicking on a document will open its content in a new tab. Users can also move back and forth in the timeline by clicking on and dragging the central portion of the window.

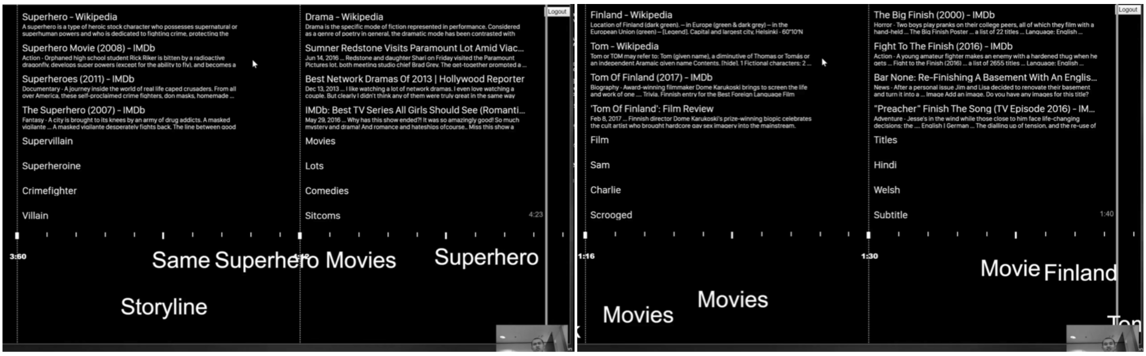
An example of system screen captures during a spoken conversation in which participants were having a natural conversation about movies is shown in Figure 3.

USER STUDY

A controlled laboratory experiment was designed to answer the research questions. SearchBot was originally conceived to support informal conversations, ideally occurring in any place. However, arranging the test in a natural environment (e.g., a cafeteria) would have subjected the system to a number of uncontrollable factors (e.g., ambient noise and incidental conversations) that could have influenced the system's performance and participants' experience. Since we did not know whether the items proposed by SearchBot could effectively feed a conversation, we chose to limit potential confounding factors by keeping the test in a controlled setting.

Experimental design

The experiment followed a within-subjects design with one independent variable being the system in use. Informal conversations on building travel or movie lists were supported either by SearchBot or by a traditional search engine used as the control. The order of presentation of systems and topics of discussion were counterbalanced across participants.



…
Participant 1: Yeah. For superhero movies I was very into it because I just-- you know when I was little boy they were all, they were all around me so…
Participant 1: That was-- that was when I thought, "Wow, a movie can be made like this".
Participant 1: Oh that's cool. The-- besides super hero movies, I'm also into dramas.
…

…
Participant 2: What's the last movie you saw in a theater?
Participant 1: [sighs] The last movie? Can-- Oh yeah, the last time I went to theater was for the movie Tom of Finland actually.
Participant 2: Oh.
I gotta see that. Probably should watch it. Was it good?
…

Figure 3. Two example screen captures of the system in a session where the participants were having a natural conversation about movies. The corresponding transcripts of the spoken conversations are shown below the screen captures. The system is recognizing and recommending entities and matching documents based on the conversational input.

Materials
We asked participants to complete three questionnaires. The first questionnaire was meant to collect background information (first and last name, age, provenance, education) and expertise in the field of HCI and previous experience with conversational agents. We also asked participants to assess the quality of the entities displayed by the system. More specifically, after they had used the experimental system, they were shown a list of the 100 suggested entities most frequently displayed during the conversation. For each word, we asked them to indicate if the word was pertinent to the conversation (namely, relevant) and if it was effectively mentioned in the conversation (namely, mentioned). A third questionnaire aimed at investigating their experience with the system in use. More specifically, we devised the questionnaire ad hoc, and it consisted of 12 items, exploring the impression that the system affected the conversation (items 1-4), the quality of the experience with the system (items 5-8), and the perceived quality of the entities shown (items 9-12). Participants had to indicate their level of agreement with each item on a 5-point Likert scale. For items 9-12, the option "not applicable" was also available. We asked participants in the experimental condition to complete all three questionnaires, and we asked participants in the control condition to complete only the third one. Finally, we devised a semi-structured interview to capture participants' direct comments and impressions of the system. More specifically, during the interview, we asked them to report their overall impressions of the system and the entities displayed. Furthermore, they had to comment on whether they had the impression that the system affected the conversation and whether they got interesting information. After the second session, we asked them to compare their experiences with both systems.

Equipment
For the present experiment, each participant used a MacBook Pro 15" laptop connected to a Samson Meteor microphone. The experimental session was video-recorded using a Panasonic camcorder. Additionally, the screen recording was taken using Screencast-O-Matic software, which also recorded participants' faces with the webcam embedded in the laptop.

Setting
The experiment took place in the laboratory. We set up the room to resemble a comfortable and informal environment, where participants could feel at ease. They sat at a desk in front of each other. Each participant had a laptop in front

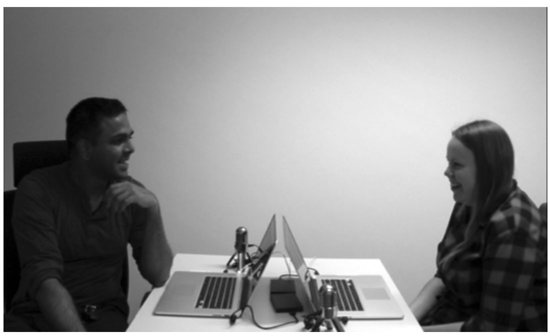


Figure 4. Experimental setup. Participants were sitting around a table, and a laptop was placed in front of each participant. The laptops were displaying the SearchBot interface. Microphones were placed on the table to record the conversation.

of him/her. The laptop allowed them to easily maintain eye contact with their conversation partners and quickly glance at the screen (Figure 4).

Procedure

The experimental procedure consisted of two main phases, each corresponding to the system in use. On the day of the test, participants were first welcomed by the experimenter, who introduced them to the experiment's main goals and overall procedure. After that, participants signed the informed consent. Phase 1 started with a training session on how to use the system; when everything was clear, the experimental session began. During the experiment, the experimenter simply asked participants to talk with their partners for 20 minutes. The experimenter then left the room to allow participants to talk freely. He followed the experimental session through a video connection and was reachable in case participants needed assistance. The task assigned was not meant to generate a specific outcome; rather, it was intended to provide only a general shape to the conversation. More specifically, we asked participants to share their experiences regarding the movies or travels (depending on the experimental condition) that had impressed them and to get inspirations from their partners' words. We did not force the participants to use the system, but we allowed them to freely utilize or ignore recommendations according to their needs. After 20 minutes, the experimenter returned to the laboratory and asked participants to complete the online questionnaire about the quality of the entities shown. Next, the experimenter accompanied one participant out of the room to complete the post-experience questionnaire while the other participant remained in the laboratory and completed the semi-structured interview. After they both finished, they swapped places. Phase 2 unfolded exactly as Phase 1 did, with the only exception being the system in use. We used the same instructions in both experimental conditions. In both cases, we left participants free to use or ignore the system according to their needs.

Control system

We used Google Custom Search to create a custom search engine that would mimic the behavior of the APIs used in the experimental condition, while maintaining interaction and the look and feel typical of traditional search engines (Figure 5).

Datasets

To provide more relevant results for our tasks, we set Google Custom Search so that it would emphasize selected websites regarding movies or travels, as well as Wikipedia. We applied the same setting to both experimental conditions, and we restricted the search engine to specified domains[3].

[3]We used the following domains in the movie task: www.hollywoodreporter.com, www.imdb.com, www.themoviedb.org, and www.rottentomatoes.com. We used the following domains in the travel task: www.wikitravel.org, www.travelandleisure.com, www.worldtravelguide.net, and www.tripexpert.com. In addition, we used the English Wikipedia for both tasks.

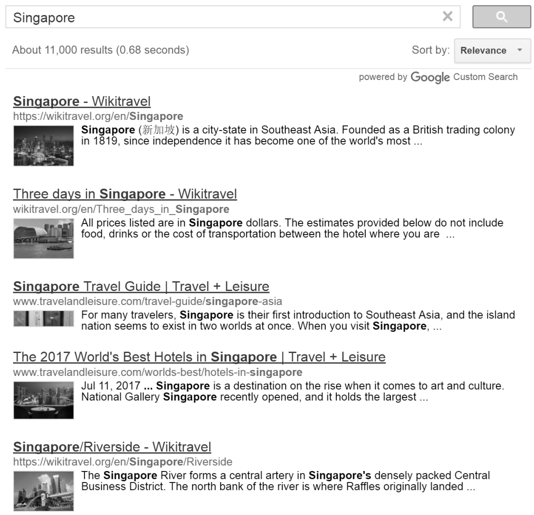


Figure 5. The instance of Google Custom Search used as a control.

Participants

A total of 24 participants (12 female) took part in the present study. The participants' mean age was 27 years ($SD = 3.87$). Of the participants, 12 were undergraduate students, five were doctoral students, three were research assistants, three were post-doc researchers, and one was a nurse. Overall, 11 participants reported having previous experience with conversational agents, and all of them reported rare usage of them. They received two movie tickets as compensation for participating in the experiment.

Measures

In order to assess the the proactive search agent' effect on the conversation, we used objective and subjective measures.

Influence of information shown on the conversation. To understand whether the information the proactive search agent presented on the screen influenced the conversation, we counted the entities extracted from the items shown on the screen that were mentioned in the 60 seconds following their first appearance on the screen. To control for possible cases in which displayed entities were mentioned by chance, we performed the same calculation in the control condition. In this case, the proactive search interface was running in the background, and results were not shown to the participants. To perform this analysis, we used a script on system logs and transcripts of the conversations obtained through a professional service.

Consumption of web resources. The number of pages opened by participants during the conversation served as a proxy for the consumption of Web resources. More specifically, the research prototype logged the pages accessed in the experimental condition, and for the control condition, the pages were traced through the navigation history of the Web browser used.

Perceived quality of the recommended items. We showed participants the list of the 100 recommended entities and the list of the 100 Web documents that the system displayed most frequently, and we asked them to mark the items that they considered pertinent and relevant to the conversation. We considered this measure a proxy for the perceived quality of the items the system suggested.

Preferred items with the proactive search agent. In the experimental condition, we logged the item types (i.e., Web documents, recommended entities, and recognized entities) that the system displayed and that the user clicked on to seek more information.

Subjective experience. We investigated participants' subjective experience with the system using a questionnaire and a semi-structured interview. We devised the questionnaire *ad hoc*, and it explored aspects related to the participants' impressions that the system had affected the conversation, the quality of the experience of use, and the overall relevance of the items displayed. The average score for each dimension was computed and then compared between the two groups. Concerning interviews, we transcribed participants' answers and ran a thematic analysis of those answers. We reviewed transcripts, identified recurring themes, and organized them into a codebook. We then applied the codes to the corpus of data [17].

FINDINGS

Influence of the information shown on the conversation

A Wilcoxon Signed-Ranks Test indicated that the entities extracted from the items shown on the screen were effectively mentioned in the conversation. More specifically, the number of extracted entities that participants mentioned in their conversations was significantly higher in the experimental condition (in which the items were actually shown) than in the control (in which the system was running in the background): $z = 2.33$, $p = .02$ (exp. $M = 7.46$, $SD = 4.90$, $Mdn = 6.00$; con. $M = 4.37$, $SD = 2.20$, $Mdn = 4.50$). This finding indicates that the references to the entities shown on the screen in the experimental condition were not due to chance (Figure 6).

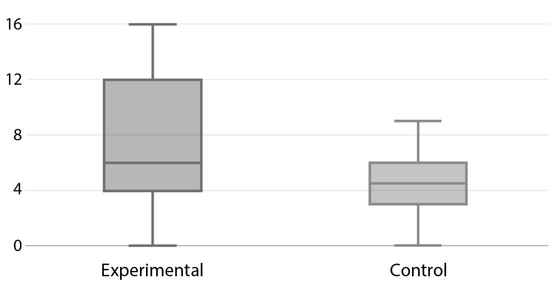


Figure 6. Number of entities that were mentioned within 60 seconds of when they were shown on the screen.

Consumption of Web resources

The consumption of Web resources, measured as the average number of Webpages opened, was also compared between the two conditions using a Wilcoxon test. The analysis did not highlight a statistical significant difference: $z = .4$, $p = .68$ (exp. $M = 11.75$, $SD = 10.01$, $Mdn = 11$; con. $M = 12.58$, $SD = 9.81$, $Mdn = 9.5$). This finding suggests that the system in use did not alter participants' search behavior.

Perceived quality of the recommended items

31.58% ($M = 31.58$, $SD = 14.70$) of the selected recommended entities were rated as relevant in the experimental condition, and the portion of the selected recommended documents rated as relevant was 17.38% ($M = 17.38$, $SD = 10.95$).

Preferred items with the proactive search agent

Figure 7 shows how participants clicked on the various types of items in the experimental condition. A non-parametric Friedman test among the various types of items clicked rendered a Chi-square value of 18.1, which was statistically significant ($p < .001$). Wilcoxon Signed-Ranks Tests with Bonferroni correction indicated that participants clicked on items representing Web documents significantly more than either suggested entities ($p < .01$) or recognized entities ($p < .01$), but there was no difference between the number of clicks on suggested entities and recognized entities ($p = .17$).

Questionnaires

We computed the average score for each dimension that the questionnaire assessed and compared the experimental and control conditions using a Wilcoxon test (Table 1). The analysis showed that participants had the impression that their conversations were affected to a greater extent by the system in the experimental condition than in the control condition: $z = 2.32$, $p = .02$ (exp. $M = 3.62$, $SD = .64$, $Mdn = 3.75$; con. $M = 3.15$, $SD = .86$, $Mdn = 3.00$). Similarly, the reported quality of the experience of use was more positive for the experimental condition ($M = 3.55$, $SD = .75$, $Mdn = 3.50$) than for the control condition ($M = 3.14$, $SD = .76$, $Mdn = 3.37$), $z = 2.01$, $p = .03$. No difference emerged between the conditions regarding the perceived relevance of the entities shown $z = .59$, $p = .55$ (exp. $M = 3.5$, $SD = .57$, $Mdn = 3.62$; con. $M = 3.4$, $SD = .51$, $Mdn = 3.5$).

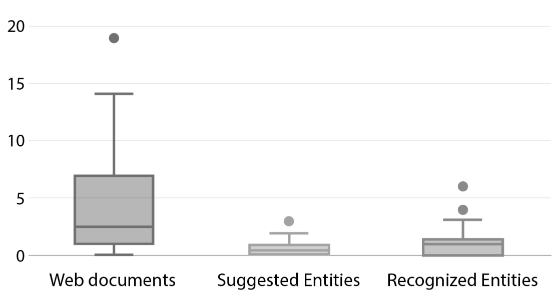


Figure 7. The types of items participants clicked on in the experimental condition. The y-axis is the number of clicks per session.

Table 1. Questionnaires

	SearchBot			Control			Comparison
	M	SD	Mdn	M	SD	Mdn	Wilc. Test
Influence on the conversation							
1. The system can give the conversation new directions	3.88	0.74	4.00	3.00	1.06	3.00	
2. The system provides little support for enriching the conversation	2.88	1.19	3.00	3.08	1.21	3.00	$\boldsymbol{z = 2.32}$
3. The system has the potential to influence what people are about to say	3.88	0.80	4.00	3.13	1.08	4.00	$\boldsymbol{p = .02}$
4. The system had almost no effect on the conversation	2.38	0.88	2.00	2.46	1.14	2.00	
Quality of the experience							
5. The system was frustrating	2.25	0.90	2.00	2.46	1.10	2.00	
6. Using the system was fun	3.67	0.87	4.00	2.88	0.85	3.00	$\boldsymbol{z = 2.01}$
7. Using the system was effortful	2.79	1.10	3.00	2.79	1.14	3.00	$\boldsymbol{p = .03}$
8. Using the system was pleasant	3.58	1.02	4.00	2.96	0.81	3.00	
Relevance of the items shown							
9. The items shown were overall interesting in the conversation	3.67	1.01	4.00	3.38	1.01	4.00	
10. The items shown were not relevant to the conversation	2.33	0.70	2.00	2.25	0.68	2.00	$z = .59$
11. The items shown were overall pertinent to the conversation	3.61	0.72	4.00	3.48	0.73	4.00	$p = .55$
12. The items shown were redundant	2.92	0.78	3.00	2.96	0.95	3.00	

Interviews

Overall, SearchBot was well-received by participants, with the majority of users reporting a positive experience ($N = 16$): *"It was fun.... It's different from everything I've tried before" (P7A)*. As interviews indicated, participants preferred SearchBot over the control condition ($N = 14$) because it allowed participants to maintain eye contact during conversation and provided users with information effortlessly:

> *I think the [experimental] system is much more useful and easier because with the [control] system, you have to make a decision–"Ok, I need to google something"–and here [experimental system], it's just a big flow, and you need to watch if something comes up. If nothing comes up, you just ignore it (P4B).*

> *The first one [experimental] helped you to keep the eye contact.... You don't need to concentrate... it gives you ideas sometimes before they come to your mind (P5A).*

> *I think this one [experimental] was better.... It was more useful. It didn't get me frustrated with wrong information, and even if it displayed the kind of stuff that were not relevant, it didn't bother my eyes. [When something relevant was shown, I thought] this is relevant to our discussion; let's click it. So it was better than the other one (P8B).*

However, some users ($N = 7$) did not convey a clear preference between the two systems although they remarked that they were different. They highlighted that the control system was more convenient to use when they explicitly wanted to look up some specific pieces of information but that using SearchBot was easier because it required no input from the participants:

> *The [experimental] one is easier because you don't have to do anything; you just say the names, and you got the links. But again, the [control] one was nice because you could use it to Google (P10B).*

A smaller proportion of participants ($N = 3$) expressed a preference for the control system because it made them feel in control:

> *[With] the [control] system, [it] was a better experience maybe because I'm used to using a search engine in my daily life all the time during a conversation (P12A).*

Notably, the majority of the participants ($N = 15$) did have the feeling that SearchBot affected the conversation either by offering the chance to deepen the current topic ($N = 8$) or by inspiring new points to discuss ($N = 7$).

> *It [the experimental system] gave us new information.... Like, we were unsure where the Red Square was, if it was Moscow or St. Petersburg, and we found that out (P7B).*

> *It [the experimental system] supported the conversation as it went. For example, we were talking about Mongolia, and I think it suggested Genghis Khan, and of course Genghis Khan is part of Mongolia (P10A).*

> *We didn't know what to talk about next, and I looked at the [experimental] system, and it said, "Berlin," so I was interested in that place, so I asked her if she had been there before, so it changed the direction (P3B).*

However, seven users did not have the impression that the experimental system was a support for the conversation, either because they felt proficient with the topic (*"I don't have that much the impression of that [...] it was a topic in which I didn't really need help [...] the topic was quite familiar, especially since we talked about places I have been" (P6A)*) or because they could not find the entities suggested by the system in a timely manner (*"Not much, to be honest ... but I was able to see the potential of it ... because it didn't catch up with what we were talking about" (P12A)*)

The majority of the participants ($N = 14$) did not find the information provided unique because they said they could have found the same data using the search engines they usually employ: *"I don't think so because I could have Googled everything it was showing" (P1B)*. Nevertheless, nine participants thought that the entities the system displayed were not obvious:

I think the [experimental] system gave me a lot of unexpected results. So, without the system, I would not search for those keywords (P3B).

SearchBot supported a better understanding of the conversations' topics, according to 16 users: *"'Cause I can quickly check something. When we were not sure what the real answer was, we could just click and check" (P3A).* However, six participants did not have the same impression because of their prior knowledge of the conversation's topic: *"Maybe not because the topic was so easy, you don't need help with this topic because you know it" (P5B).*

The entities displayed by the system were generally considered either relevant or useful ($N = 14$):

We talked about the recent Star Wars *movie, and the system displayed links to the relevant pages. We were discussing some details, and we were able to check them from the pages that were proposed (P8A).*

They [the recommendations] were pretty good . . . When you say something, you have four to five different choices, and I think they were pretty accurate. Sometimes even you say something not so important, you would find something important. For example, we talked about cheap, how cheap is gonna be, and instantly the system gave us, like, cheap flights (P7A).

When it was relevant, it was really useful because you just click on it and then find more information, for example the year [of the movie] or something like that (P9A).

However, three users found the entities not relevant for the conversation and thus distracting: *"Something totally not relevant appeared, which made me less concentrated on the conversation" (P9B).* Two participants commented that the search based on the combination of two entities didn't work for them: *"I was talking about how to go to Bordeaux by train, so it was [the system suggested] train and Bordeaux separately, but it would be much more useful if it was like 'train to Bordeaux'" (P6A).* Three participants complained that the entities were displayed too late with respect to the stream of conversation: *"Maybe it showed it [the entities] a little bit afterward ... but it was interesting information" (P7B).* Finally, only one participant reported that the entities were useless to the conversation.

DISCUSSION

Information spaces grow in size and richness, and users increasingly prefer information to be delivered to them proactively as a part of secondary tasks supporting their primary tasks. Consequently, conventional search interfaces fall short in allowing users to concentrate on their primary tasks, and supporting information access by anticipating users' needs has become a major bottleneck in many complex tasks.

We studied how proactive searches conducted by using input directly from natural conversations between individuals can support the conversations. We designed the SearchBot system and used it in an experiment to study the influence, the number of consumed resources, and the effect on the user experience of a proactive search interface in supporting conversations.

Answers to the Research Questions

Here we reflect on the research questions that we defined earlier.

RQ1: Does proactive search with spoken input from natural conversation influence the conversation? Yes. Figure 6 shows how participants in the experimental condition frequently referred to the entities and documents shown on the screen during their conversations. The comparison with the control condition, in particular, demonstrated that these references were not due to chance. This result indicates that not only did the proactive search system retrieve useful information, but the displayed information influenced the conversation, as questionnaires and interviews further confirmed.

RQ2: Does proactive search with spoken input from natural conversation affect the consumption of Web resources during conversations? No. There was no significant difference in the number of Web resources consulted between the experimental and control systems. This result suggests that participants retrieved the same number of useful resources supporting the conversation in both experimental conditions. However, while in one case the resources were automatically retrieved by the proactive search agent, in the other case, explicit query formulation and refinement was needed.

RQ3: Does proactive search with spoken input from natural conversation affect participants' subjective experience? Yes. In general, the reported quality of the experience of using the system was more positive for the experimental condition, as it allowed participants to keep eye contact with each other, enabling more fluent conversation. Participants reported that SearchBot allowed them to check facts and build common ground without needing to exert much mental effort. Furthermore, the system was able to expand the conversation in new directions. However, the added value of the proactive search experience seemed to come with the cost of feeling less in control of the search process. All in all, the participants were more satisfied with the SearchBot system.

Design Implications

In this section, we start from the lessons we learned in our study and discuss design implications to help set the stage for future developments of proactive search interfaces for conversation support.

Relevance of recommendations in conversation support. Our study proved that errors in recognition and the consequent display of non-relevant results do not necessarily prevent a proactive search system from effectively supporting a conversation. Most participants were able to easily ignore non-relevant recommendations while benefiting from the relevant ones. However, while users can easily skim through a screen full of non-relevant results to identify a single piece of relevant information, it is important that at least that single piece of relevant information is there when needed. When this didn't happen, participants experienced distraction, frustration, and loss of trust in the system. Some heuristics may be required to improve the relevance of displayed results. In our case, we mostly used queries involving all the entities contained in a recognized sentence. However, this did not always produce the desired results. Therefore, we used knowledge gained from pilot experiments to prioritize locations and people as the entities particularly relevant to our tasks. When such entities were detected, we used them in single-entity queries and showed the results on top of the list. This generally improved the system's capability to support the conversations, but it came at the cost of showing fewer recommendations deriving from combined entities. Deciding how to combine search terms and if and how to prioritize special terms is a key aspect to be considered when designing systems for proactive search support in conversations.

Types and number of recommended items. Our findings suggest that with SearchBot, around one third of recommended entities and around 17% of recommended documents were relevant to the conversation. Nevertheless, our findings also show that the most used items were Web documents. While in this initial investigation we chose to display the same number of recommended entities and documents, our findings suggest that future implementations should carefully consider how to allocate the screen's real estate for various kinds of items according to the type of discussion to be supported. In our study, participants did not use recommended entities much. They used them mostly to expand the conversation with new ideas. While this result confirms past research on creative conversations [4], it also highlights the fact that various item types are needed in conversations that are not merely creative, such as those explored in our study. Our findings suggest that richer sources of information, such as Web documents, should be given more importance in those cases, as they allow participants to check facts and build a common understanding on the topic of discussion.

Combining proactive and explicit searches. Our study suggests that proactive searches performed automatically by the system by using content extracted from spoken conversation allowed users to more easily maintain eye contact and stay focused on the conversation, as the interviews indicated. Most people preferred the proactive search approach to the explicit search one, but this preference came at the cost of losing control over the system. As the two approaches showed complementary strengths, next developments should consider integrating both modes of operations in the same interface.

Limitations and Future Work

While our work shows that proactive search support in conversation is already possible and provides several advantages over relying on traditional explicit search, it also has some limitations. Even if the system was devised to support informal and casual conversation, we chose to test it in a controlled setting. While this arrangement allowed us to control for confounding factors, it also limited our findings' ecological validity. Further research is needed to understand how proactive search support can affect conversation in more natural settings. Also, in this work, we designed a prototype with a limited set of features. The prototype served as a research tool to study the potential of proactive searches from natural spoken input to support conversations. This strategy allowed us to better understand the different features' roles. However, to understand the real potential of proactive search support in conversations, future implementations should consider integrating more sophisticated features, such as interactions with the system to build and maintain intent models [5, 37] and using topic modeling to extract relevant context and improve the relevance of retrieved results.

CONCLUSION

The approach to proactive search used in this work utilizes the subtle human feedback signals observed directly from natural conversations, as opposed to previous work, which has mainly relied on conventional user input, such as issued queries or visited documents. We investigated how a proactive search agent that uses vocal conversational input could support informal conversations on travel or movie lists. We designed SearchBot, a proactive search agent that listens to conversations, detects entities mentioned in the conversations, and proactively retrieves and presents information related to the conversations. We used SearchBot in a comparative study with 12 pairs of participants. Our findings showed that information retrieved proactively by an agent listening to the conversation had the potential to effectively support the conversation with facts and ideas without causing much interruption to the conversation's flow but at the cost of participants feeling less in control of the search process. Findings also show that the proactive search approach retrieved the same number of useful resources supporting the conversation but without the participants needing to formulate explicit queries. Notably, this study allowed us to explore the design space of proactive search support in conversations, providing key design implications for the paradigm's future developments.

ACKNOWLEDGMENTS

This research was partially funded by the Academy of Finland (286440, 312274, 313610).

REFERENCES

1. Jesper Aagaard. 2016. Mobile devices, interaction, and distraction: a qualitative exploration of absent presence. *AI & SOCIETY* 31, 2 (01 May 2016), 223–231. DOI: http://dx.doi.org/10.1007/s00146-015-0638-z

2. Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 6 (June 2005), 734–749. DOI: http://dx.doi.org/10.1109/TKDE.2005.99

3. Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving Web Search Ranking by Incorporating User Behavior Information. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '06)*. ACM, New York, NY, USA, 19–26. DOI: http://dx.doi.org/10.1145/1148170.1148177

4. Salvatore Andolina, Khalil Klouche, Diogo Cabral, Tuukka Ruotsalo, and Giulio Jacucci. 2015a. InspirationWall: Supporting Idea Generation Through Automatic Information Exploration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and*

Cognition (C&C '15). ACM, New York, NY, USA, 103–106. DOI: http://dx.doi.org/10.1145/2757226.2757252

5. Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. 2015b. IntentStreams: Smart Parallel Search Streams for Branching Exploratory Search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 300–305. http://doi.acm.org/10.1145/2678025.2701401

6. Salvatore Andolina, Hendrik Schneider, Joel Chan, Khalil Klouche, Giulio Jacucci, and Steven Dow. 2017. Crowdboard: Augmenting In-Person Idea Generation with Real-Time Crowds. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition (C&C '17)*. ACM, New York, NY, USA, 106–118. DOI: http://dx.doi.org/10.1145/3059454.3059477

7. Tony Bergstrom and Karrie Karahalios. 2009. Conversation Clusters: Grouping Conversation Topics Through Human-computer Dialog. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 2349–2352. DOI: http://dx.doi.org/10.1145/1518701.1519060

8. Dan Bohus and Eric Horvitz. 2009. Dialog in the Open World: Platform and Applications. In *Proceedings of the 2009 International Conference on Multimodal Interfaces (ICMI-MLMI '09)*. ACM, New York, NY, USA, 31–38. DOI:http://dx.doi.org/10.1145/1647314.1647323

9. Barry Brown, Moira McGregor, and Donald McMillan. 2015. Searchable Objects: Search in Everyday Conversation. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 508–517. DOI: http://dx.doi.org/10.1145/2675133.2675206

10. Huanhuan Cao, Derek Hao Hu, Dou Shen, Daxin Jiang, Jian-Tao Sun, Enhong Chen, and Qiang Yang. 2009. Context-aware Query Classification. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*. ACM, New York, NY, USA, 3–10. DOI: http://dx.doi.org/10.1145/1571941.1571945

11. David Carmel, Naama Zwerdling, Ido Guy, Shila Ofek-Koifman, Nadav Har'el, Inbal Ronen, Erel Uziel, Sivan Yogev, and Sergey Chernov. 2009. Personalized Social Search Based on the User's Social Network. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*. ACM, New York, NY, USA, 1227–1236. DOI: http://dx.doi.org/10.1145/1645953.1646109

12. Zhicong Cheng, Bin Gao, and Tie-Yan Liu. 2010. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th international conference on World wide web*. ACM, 221–230.

13. Anind K. Dey, Gerd Kortuem, David R. Morse, and Albrecht Schmidt. 2001. Situated Interaction and Context-Aware Computing. *Personal Ubiquitous Comput.* 5, 1 (Jan. 2001), 1–3. DOI: http://dx.doi.org/10.1007/PL00000013

14. Joan Morris DiMicco and Walter Bender. 2004. Second messenger: increasing the visibility of minority viewpoints with a face-to-face collaboration tool. In *Proceedings of the 9th international conference on Intelligent user interfaces*. ACM, 232–234.

15. Joan Morris DiMicco, Anna Pandolfo, and Walter Bender. 2004. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 614–623.

16. Desmond Elliott and Joemon M. Jose. 2009. A Proactive Personalised Retrieval System. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*. ACM, New York, NY, USA, 1935–1938. DOI: http://dx.doi.org/10.1145/1645953.1646269

17. Greg Guest. 2014. *Public health research methods*. Sage Publications.

18. Ramanathan Guha, Vineet Gupta, Vivek Raghunathan, and Ramakrishnan Srikant. 2015. User Modeling for a Personal Assistant. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*. ACM, New York, NY, USA, 275–284. DOI: http://dx.doi.org/10.1145/2684822.2685309

19. Qi Guo and Eugene Agichtein. 2010. Ready to Buy or Just Browsing?: Detecting Web Searcher Goals from Interaction Data. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*. ACM, New York, NY, USA, 130–137. DOI: http://dx.doi.org/10.1145/1835449.1835473

20. Monika Henzinger, Bay-Wei Chang, Brian Milch, and Sergey Brin. 2005. Query-Free News Search. *World Wide Web* 8, 2 (2005), 101–126. DOI: http://dx.doi.org/10.1007/s11280-004-4870-6

21. Jonathan Kilgour, Jean Carletta, and Steve Renals. 2010. The Ambient Spotlight: Queryless Desktop Search from Meeting Speech. In *Proceedings of the 2010 International Workshop on Searching Spontaneous Conversational Speech (SSCS '10)*. ACM, New York, NY, USA, 49–52. DOI: http://dx.doi.org/10.1145/1878101.1878112

22. David Kirsh. 2014. The importance of chance and interactivity in creativity. *Pragmatics & Cognition* 22, 1 (2014), 5–26.

23. Weize Kong, Rui Li, Jie Luo, Aston Zhang, Yi Chang, and James Allan. 2015. Predicting Search Intent Based on Pre-Search Context. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '15)*. ACM, New York, NY, USA, 503–512. DOI: http://dx.doi.org/10.1145/2766462.2767757

24. Henry Lieberman. 1995. Letizia: An agent that assists web browsing. In *Proc. IJCAI (1)*. 924–929.

25. Daniel J. Liebling, Paul N. Bennett, and Ryen W. White. 2012. Anticipatory Search: Using Context to Initiate Search. In *Proc. SIGIR*. ACM, 1035–1036.

26. Moira McGregor and John C. Tang. 2017. More to Meetings: Challenges in Using Speech-Based Technology to Support Meetings. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 2208–2220. DOI: http://dx.doi.org/10.1145/2998181.2998335

27. Donald McMillan, Antoine Loriette, and Barry Brown. 2015. Repurposing Conversation: Experiments with the Continuous Speech Stream. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3953–3962. DOI: http://dx.doi.org/10.1145/2702123.2702532

28. Michael F McTear. 2002. Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys (CSUR)* 34, 1 (2002), 90–169.

29. Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781 (2013). http://arxiv.org/abs/1301.3781

30. Takashi Nakamura. 2015. The action of looking at a mobile phone display as nonverbal behavior/communication: A theoretical perspective. *Computers in Human Behavior* 43 (2015), 68 – 75. https://doi.org/10.1016/j.chb.2014.10.042

31. Matteo Negri, Marco Turchi, José G. C. de Souza, and Daniele Falavigna. 2014. Quality Estimation for Automatic Speech Recognition. In *COLING 2014, 25th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, August 23-29, 2014, Dublin, Ireland*. 1813–1823. http://aclweb.org/anthology/C/C14/C14-1171.pdf

32. Martin Porcheron, Joel E. Fischer, and Sarah Sharples. 2016. Using Mobile Phones in Pub Talk. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 1649–1661. DOI: http://dx.doi.org/10.1145/2818048.2820014

33. Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 22, 4 (01 Oct 2012), 317–355. DOI: http://dx.doi.org/10.1007/s11257-011-9115-7

34. Paul Resnick and Hal R. Varian. 1997. Recommender Systems. *Commun. ACM* 40, 3 (March 1997), 56–58. DOI: http://dx.doi.org/10.1145/245108.245121

35. Bradley Rhodes and Thad Starner. 1996. Remembrance Agent: A continuously running automated information retrieval system. In *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology*. 487–495.

36. B. J. Rhodes and P. Maes. 2000. Just-in-time Information Retrieval Agents. *IBM Syst. J.* 39, 3-4 (July 2000), 685–704. DOI: http://dx.doi.org/10.1147/sj.393.0685

37. Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. 2014. Interactive Intent Modeling: Information Discovery Beyond Search. *Commun. ACM* 58, 1 (Dec. 2014), 86–92. DOI: http://dx.doi.org/10.1145/2656334

38. G. Salton, A. Wong, and C. S. Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (Nov. 1975), 613–620. DOI: http://dx.doi.org/10.1145/361219.361220

39. Gianluca Schiavo, Alessandro Cappelletti, Eleonora Mencarini, Oliviero Stock, and Massimo Zancanaro. 2014. Overt or subtle? Supporting group conversations with automatically targeted directives. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 225–234.

40. Yang Shi, Yang Wang, Ye Qi, John Chen, Xiaoyao Xu, and Kwan-Liu Ma. 2017. IdeaWall: Improving Creative Collaboration Through Combinatorial Visual Stimuli. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 594–603. DOI: http://dx.doi.org/10.1145/2998181.2998208

41. Sosuke Shiga, Hideo Joho, Roi Blanco, Johanne R. Trippas, and Mark Sanderson. 2017. Modelling Information Needs in Collaborative Search Conversations. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 715–724. DOI: http://dx.doi.org/10.1145/3077136.3080787

42. Yang Song and Qi Guo. 2016. Query-Less: Predicting Task Repetition for NextGen Proactive Search and Recommendation Engines. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 543–553. DOI: http://dx.doi.org/10.1145/2872427.2883020

43. Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2005. Personalizing Search via Automated Analysis of Interests and Activities. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*. ACM, New York, NY, USA, 449–456. DOI: http://dx.doi.org/10.1145/1076034.1076111

44. Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2010. Potential for Personalization. *ACM Trans. Comput.-Hum. Interact.* 17, 1, Article 4 (April 2010), 31 pages. DOI: http://dx.doi.org/10.1145/1721831.1721835

45. Jaime Teevan, Amy Karlson, Shahriyar Amini, A. J. Bernheim Brush, and John Krumm. 2011. Understanding the Importance of Location, Time, and People in Mobile Local Search Behavior. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 77–80. DOI: http://dx.doi.org/10.1145/2037373.2037386

46. Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017a. Proactive Information Retrieval via Screen Surveillance. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 1313–1316. DOI: http://dx.doi.org/10.1145/3077136.3084151

47. Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017b. Watching Inside the Screen: Digital Activity Monitoring for Task Recognition and Proactive Information Retrieval. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 109 (Sept. 2017), 23 pages. DOI: http://dx.doi.org/10.1145/3130974

48. Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers, [San Rafael, Calif.].

49. Biao Xiang, Daxin Jiang, Jian Pei, Xiaohui Sun, Enhong Chen, and Hang Li. 2010. Context-aware ranking in web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 451–458.

Publication VII

Salvatore Andolina, Pedram Daee, Tung Vuong, Tuukka Ruotsalo, Khalil Klouche, Mats Sjöberg, Samuel Kaski,and Giulio Jacucci. Proactive Recommendation in Context: From Relevant Items to Actionable Entities. Submitted to *ACM Transactions on Computer-Human Interaction*, February 2019.

Proactive Recommendation in Context: From Relevant Items to Actionable Entities

SALVATORE ANDOLINA, Aalto University, Finland
PEDRAM DAEE, Aalto University, Finland
TUNG VUONG, University of Helsinki, Finland
TUUKKA RUOTSALO, Aalto University, Finland
KHALIL KLOUCHE, University of Helsinki, Finland
MATS SJÖBERG, Aalto University, Finland
SAMUEL KASKI, Aalto University, Finland
GIULIO JACUCCI, University of Helsinki, Finland

Proactive recommendation systems aim to support users in their everyday digital life by automatically retrieving information that can be useful for the task at hand. However, fairly little attention has been devoted to the design of proactive recommendation approaches for everyday digital tasks enabling retrieval of actionable entities across application boundaries. We present the design and implementation of an entity-centric proactive recommendation system that makes recommendations by capturing user's digital context. The context consist of entities, such as people, documents, applications, and keywords, detected via digital screen monitoring from everyday digital tasks. The entities are used to build task-related context structures that, in turn, are used to recommend actionable entities that users can utilize for initiating new digital activity. We investigate whether the entity-centric approach can effectively support real-world everyday digital tasks by providing recommendations that are relevant and *useful*, i.e., have a concrete influence on users' tasks. Quantitative and qualitative results from a study with 13 participants resuming their own real-world heterogeneous tasks demonstrate the relevance and usefulness of recommendations. Notably, our analysis of entity recall shows that the system allowed users to find more relevant entities than they would find on their own. Furthermore, quantitative measures of influence show that, in experimental condition, about 21% of documents and 29% of applications participants used to perform their tasks would not have been used without the system recommendation. The proposed system and study demonstrate the importance of proactive entity recommendation to enable effortless access to information, promote effective use of relevance feedback, re-establish lost connections, recall useful resources that have been forgotten, and get timely recommendations of actionable entities in a variety of everyday digital tasks.

CCS Concepts: • **Information systems → Users and interactive retrieval**; **Recommender systems**; • **Human-centered computing → Human computer interaction (HCI)**;

Additional Key Words and Phrases: Proactive search, user intent modeling

Salvatore Andolina, Pedram Daee, and Tung Vuong contributed equally.
Authors' addresses: Salvatore Andolina, Aalto University, Helsinki, Finland, salvatore.andolina@aalto.fi; Pedram Daee, Aalto University, Helsinki, Finland, pedram.daee@aalto.fi; Tung Vuong, University of Helsinki, Helsinki, Finland, vuong@cs.helsinki.fi; Tuukka Ruotsalo, Aalto University, Helsinki, Finland, tuukka.ruotsalo@helsinki.fi; Khalil Klouche, University of Helsinki, Helsinki, Finland, khalil.klouche@helsinki.fi; Mats Sjöberg, Aalto University, Helsinki, Finland, mats.sjoberg@aalto.fi; Samuel Kaski, Aalto University, Helsinki, Finland, samuel.kaski@aalto.fi; Giulio Jacucci, University of Helsinki, Helsinki, Finland, giulio.jacucci@helsinki.fi.

1073-0516/2019/2-ART0 $15.00
https://doi.org/0000001.0000001

Fig. 1. An example graphic design task for a manuscript. The system continuously extracts information from the user's screen (e.g., entities EEG, STIM, ET, Adobe Illustrator, ...), discovers the user's evolving intent, and proactively provides real-time entity recommendations (e.g., the related manuscript, other related vector graphic files, or co-authors to contact and seek feedbacks on the figure) that could help with the task.

ACM Reference Format:
Salvatore Andolina, Pedram Daee, Tung Vuong, Tuukka Ruotsalo, Khalil Klouche, Mats Sjöberg, Samuel Kaski, and Giulio Jacucci. 2019. Proactive Recommendation in Context: From Relevant Items to Actionable Entities. *ACM Trans. Comput.-Hum. Interact.* 0, 0, Article 0 (February 2019), 34 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Information retrieval (IR) is an important activity that pervades our everyday digital life. Whether we plan a vacation, write an essay, or engage in an online discussion with friends, we often need to retrieve additional information to support our tasks. However, information retrieval systems are not tailored to support peoples' digital tasks as such, but require explicit commands, input of preferences and queries, and human attention to guide the process, which take users' focus and effort away from the primary tasks. Furthermore, traditional systems do not typically address the problem of users not knowing what they do not know [48].

Proactive recommendation systems can address these challenges by continuously watching a person's context and presenting information that may be useful without requiring any action on the part of the user [43]. Available systems have been used to automatically retrieve potentially useful resources in a variety of situations, ranging from watching news [23] to writing documents [30, 42] and emails [15], or arranging meetings [60]. These systems have focused largely on the problem of providing relevant recommendations and have all been successful to some extent. Yet, relevance is only a partial measure of success: the recommended resources have a concrete impact on people's tasks only when they are used.

Understanding to what extent proactive recommendations may be relevant and *useful* in everyday digital tasks is a problem that has been less explored. There are three main challenges that have made it difficult to design a recommendation system that could provide value to users in a variety of real world everyday digital tasks and could thus be used to more rigorously study the relevance and usefulness of proactive recommendations:

- *Limited access to entities.* Everyday digital tasks require users to perform actions on several kinds of information entities [48], such as contacting people, opening applications, and so on, but current proactive recommendation systems typically are able to recommend only documents or keywords, which expose less affordances to the users, therefore providing them with less practical value.

- *Incomplete access to task context.* Everyday digital tasks often require access to information that is fragmented across several applications and services [5], but current proactive recommendation systems typically only have access to partial data which is only obtainable through predefined applications or services.
- *Handling interactivity of tasks.* Complex search tasks are highly interactive [26, 36], but current proactive recommendation systems typically do not included much support for interactivity [7, 31, 33, 42, 60]. Including features for interactive relevance feedback [46] could improve the overall quality of recommendations [44, 46], but people may not always be willing to provide explicit relevance assessments [18, 58] if the mechanism is too orthogonal to their current task. The challenge is finding efficient ways to combine implicit and explicit feedback, providing users with the right affordances for giving explicit feedback when needed.

Entity-centric modeling approaches can extract meaning from the user activity and can thus be used to provide users with more meaningful and diverse recommendations. Recent studies also suggest they foster interactivity and exploration in complex search tasks [1, 2, 29]. However, many modeling approaches (e.g. [32, 40]) are based on logs of interaction events, e.g. the use of a file, the browsing of a web page, or the execution of an application, which have limited access to entities apart from file's metadata. Content-based approaches such as those that leverage the written content of documents [15, 30, 31] are better suitable to entity extraction as they are based on richer data. However, they have typically relied on simple term extraction, leaving the opportunity to use richer sources of context unexplored. Furthermore, current content-based approaches have often been tailored to specific applications, such as for example browsers, mail clients, or word processing software, and although they had the potential to be extended, new wrappers needed to be written to analyze context from other applications.

We propose a novel approach for proactive recommendations during everyday digital tasks that overcomes the limitations of existing research by being entity-centric, interactive, and application independent. The solution we employ uses digital activity monitoring capable of extracting the texts shown on people's computer screens across application boundaries. We use this rich source of context to extract entities such as documents, applications, people, and topics during human-computer interaction. We implement a semi-supervised intent learning algorithm to learn the user interest over all entities, proactively, by monitoring user digital activity inside the screen, and/or interactively by explicit feedback. The method exploits the information in thousands of screen frames collected from the user's screen to detect the context and to recommend novel items related to the current user task. By being able to extract entities across applications boundaries, the approach is able to operate a *holistic* user modeling of intent considering relations between entities in evolving task contexts.

We evaluate the approach through a study using participants' real-life data and tasks. The system is installed on a user's laptop for two weeks for unsupervised learning of representation of entities and their relationship during actual work tasks. After this users participate in an evaluation session resuming previous work tasks. The system is set up as a separate screen where recommended entities are visualized during users' work. Users may open recommended documents, applications, and contacts or give a feedback by selecting an entity which in turn updates the set of recommended entities. This system setup is illustrated in Figure 1. A control condition is used to better investigate the benefit of the system and verify, without visualization and user feedback, to which extent predicted entities are relevant. The user study demonstrates the viability of the approach by showing the added value provided by the system in terms of discovery of novel and relevant entities. More importantly, results indicate that recommended entities were found useful and influenced the user tasks, leading to improved user experience in completing the task. The contributed approach bear as important implications the potential of entity based user interfaces as well as the opportunity to model the evolving intent and context of users across application boundaries.

2 BACKGROUND AND RELATED WORK

Recommendation systems are increasingly affecting our everyday life, due to their capability to help find relevant information in our fast-expanding digital universe. They monitor contextual signals to create and update a user profile, which is then used to recommend information tailored to the user's context (e.g., [22, 35]). This work focuses on recommendation systems that are *proactive*, i.e., those that don't require users to perform any specific action but leverage users' context and past interactions to anticipate users' needs [32] and provide them with information that is likely to be relevant in an automatic way. In particular, our work is related to *just-in-time* proactive recommendation systems [43], those that attempt to retrieve and deliver information when the user is most likely to need it.

Table 1 summarizes several attempts to create proactive information retrieval support for a variety of tasks varying from conventional web search [16, 31, 32] to document writing and meeting preparation [15, 30, 60]. They mainly differ in their capability to capture context, in the type of recommended items, and in the extent to which they allow users to explicitly affect their models through interactive feedback.

This prior work, however, has mainly focused on how relevant document recommendations could be provided within tasks that were either simulated [16, 30], or otherwise limited in scope [43, 60], failing to provide a clear picture on the more interesting problem of how proactive recommendations *influence* real-world tasks. In contrast, we aim to design a proactive recommendation system that would be able to support people in a wide variety of everyday digital tasks, which, in turn, would permit to evaluate with more rigor the influence and usefulness of proactive recommendations on people's real-world tasks.

In this section, we first review insights originating from earlier research on proactive information retrieval systems with a special focus on those supporting primary tasks. Then, we analyze how previous approaches have considered the opportunity to use entities and interactivity. Finally, we review the notion of digital activity monitoring and delve into how it can be leveraged to create recommendation systems that model users more holistically by capturing most of the human-computer interaction.

2.1 Supporting Primary Tasks with Proactive Recommendations

Recent trends in information seeking research promote a view of the search activity as something belonging to a wider high-level task [24]. As a result, a large body of research has started investigated novel ways to model the context of users and use such models to infer users' intents. This has led, for example, to the emergence of paradigms such as *search personalization*, where the user's task or search context is modelled [4, 51, 53] using previous queries and page visits [56], or click-through data [8], or by modeling the task that motivates the information need [28, 37].

Proactive search [16] (or *anticipatory search* [32]) is a natural extension of search personalization with the explicit query step removed. Here the user's context is continuously being monitored and the user model updated in order to anticipate the upcoming information need. An early example of this paradigm is the *Remembrance Agent* [42] which monitors a user's personal data, e.g., emails and text documents and continuously displays a list of documents related to what the user is doing now. Here document relevance is simply estimated based on the frequency of common words with the currently active text, there is no long-term modeling. *Letizia* [31] is a similar early example

Table 1. Comparison of proactive recommendation systems

	Recommended Items	Context Extraction	Interactive Feedback	Just-in-time	Activity supported	Evaluation of usefulness
Letizia [31]	Web documents	Browsing behavior	No	Yes	Web search	No
Remem-brance Agent [42]	Documents	Emails and written notes	No	Yes	Writing a newspaper-style article [43]	Subjective scores
Elliot and Jose [16]	Documents	Browser	Yes	Yes	Multi-session search	No
Koskela et al. [30]	Web documents	Text from text editor	On keywords	Yes	Document writing	Number of selected recommendations
Watson [7]	Web documents and images	Text from text editor and browser	No	Yes	Document writing and web search	Potential usefulness of first recommendation (subjective scores)
SidePoint [33]	Text snippets and images	Text from presentation authoring software	No	Yes	Presentation writing	Qualitative feedback on the system (lab study)
CAPERS [60]	Emails	Calendar	No	Yes	Meeting preparation in an enterprise	Number of clicking and hovering actions (field experiment)
IQ [15]	Documents, people, topics	Emails	On overall quality of results	Yes	Reading or composing emails	No
CAAD [40]	Documents, applications, email addresses	Apps that make native OS calls	Update clusters of context structures	Yes	Computer-based information work	Perceived usefulness from questionnaires and interviews (field experiment)
Vuong et al. [55]	Documents and keywords	Any app	No	No	Task detection and proactive retrieval within a closed set of tasks	No
Our system	Entities (documents, people, applications, topics)	Any app	One-click explicit feedback on entities	Yes	Everyday digital tasks	Influence of recommendation on tasks plus interviews (field data collection followed by a lab phase)

that provides automatic recommendations during web browsing. A more modern application can be found in current smart phones, for example in the form of *Google Now*, which tries to model not only short-term search intents, but also long-term interests and habits based on several months of collected data [20]. Here, user-specific context classes (e.g., tasks, interests, or habits) are identified from their search history. Another approach is to extract patterns related to the time of day [50], for example that a certain task is usually done in the morning, and use these to anticipate resources that the user will need at that time.

Proactive recommendations while performing specific tasks have also been studied. For example, in [7], contextual text and image queries are performed based on text written by the user in a word processing application, and, in [34], reference recommendations are shown in a similar scenario. Other examples include proactively showing personal documents related to the current email being read or written by the user [15], or during authoring of a Powerpoint presentation [33]. In [30], a more generic approach to capturing the search intent from the primary task context is proposed. However, the experimental part mainly studies the writing task.

In contrast to these earlier approaches, we propose a more general system based on screen recording, which is not restricted to a specific task or application. In addition to dramatically expanding its scope of usage, this also facilitates collecting data across all computer activities and thus enabling more complete modeling of the entire human-computer interaction.

Contrary to other approaches, we aim to build a system that could be used to study relevance and influence of proactive recommendation systems in supporting a variety of heterogeneous everyday digital tasks. Prior studies have pointed out the need to more rigorously evaluate usefulness of recommended information [7, 43], as relevance does not necessarily correlate with usefulness. In [7], to determine whether or not the sources returned by Watson were useful in the context of a particular task, authors asked six participants to send them a copy of the last paper they wrote, fed the paper to Watson, and returned the first list of recommendations retrieved by the system back to the participants for subjective assessment of usefulness. Similarly, qualitative evaluations of usefulness have been conducted in Remembrance Agents, through subjective scores given by participants [43], and in SidePoint, through qualitative feedback on the system [33]. Other more quantitative approaches have considered interaction events such as number of clicks on recommended items as proxies of usefulness [30, 60]. Although these studies reveal potential benefits and challenges of proactive information retrieval, they fail to provide a realistic picture of the actual influence of proactive recommendation on everyday digital tasks, because of their reliance on simulated work tasks and their limited scope. The study that goes closer to our vision is presented in [40] with the evaluation of the CAAD system. CAAD automatically generates task representations (as context structures) from logs of low-level interaction events. Contrary to most systems, CAAD captures context from most applications, i.e., those that make a native OS call. The evaluation of the system was conducted in the working environment of participants over three days. Questionnaires of perceived usefulness and interviews were used to gather qualitative feedback on the system. One limitation of the study was that some participants used applications that didn't make any native OS call and were therefore invisible to CAAD. Although the study used real-world tasks, it shared most of the limitations of other studies, including the limited insights provided by subjective scores of usefulness and the overall focus on perceived accuracy of suggestions. Other limitations of all these studies include the limited usage of semantics in the provided recommendations, which might have limited their perceived usefulness, and the limited use of interactive feedback, which might have mitigated the typical loss of control experienced by people using proactive recommender systems.

In contrast, we study how proactive recommendations can be useful in everyday digital tasks by measuring their actual influence on people's task both quantitatively and qualitatively through an in-the-wild data collection followed by a lab phase where people resumed their real-world tasks. Contrary to other studies, we also employ a system that can provide more meaningful and actionable recommendations through an entity-centric approach and the use of interactive feedback. As typical everyday digital tasks require users to switch across several applications, our system is not confined to specific applications. By employing screen recording, our system is able to model task-related context comprehensively, avoiding compatibility issues experienced in prior research [40].

2.2 Entity-Based Interactions in Information Retrieval

In information retrieval, entities are references to real-world objects or concepts (e.g., persons, places, movies, topics, or products). In web searches, most emitted queries pivot around a specific entity [39]. This can be seen as a generalization of the keyword concept used by previous systems, but with clearer affordances in the computer interface (e.g. clicking on a people entity could initiate an action to contact that person).

Entity-based queries in current search engines result not only in a relevant entry (e.g., from Wikipedia) but in a knowledge graph with relevant information and related entities, providing quick links to further the exploration. Miliaraki et al. [38] studied the behavior of users of Yahoo Spark, a system that recommends related entities alongside Yahoo Search results; the users take advantage of the system to engage in exploratory entity search by discovering information through successive clicks on recommended entities. Recent research work explores novel interaction techniques through direct manipulation of displayed entities [1, 29, 45]. These systems display results as interactive objects that can be used as a query or part of a query in a new search. Benefits of these entity-based approaches come mainly from their reliance on users having to interact with information they recognize (e.g. recommended or retrieved entities) over users having to recall information, for example when typing a query. When providing both options, entity-based systems have been shown to substantially reduce the need for typing, which make them especially useful for touch devices [29].

In this paper we propose using entities as the basic unit for representing the user's evolving intent. Unlike other approaches for intent visualization and manipulation mainly based keywords [17], we use entities because they provide a richer source of information and allow us to learn more expressive user models. Entities represent familiar real-world objects and concepts, and could thus, as interactive objects, provide the right affordances for letting the users interactively refine their intent models.

2.3 Digital Activity Monitoring

Research on digital activity monitoring and prediction of user behavior has typically focused on large-scale tracking, e.g., based on what people are sharing on social media [59, 61]. This mass monitoring approach has some important draw-backs, including loss of privacy and lack of trust for the system [9]. However, some recent work has studied technologies for individual monitoring of personal data, putting the collection and analysis of the data into the hands of the individuals themselves [12, 49]. The work described in this paper completes the picture by providing the modeling and user interaction for enabling this vision.

Most of the approaches mentioned so far have focused on monitoring specific applications or other limited data sources. However, recent work [54, 55] has explored using screen monitoring, which captures the entire visual content of the computer screen for task recognition. Latent Semantic Analysis [13] with a simple bag-of-words data representation was found to be the most effective to detect users' tasks and helpful for proactive information retrieval. This tracking "inside the screen" paradigm has the benefit of being more general, as any visually communicated information can potentially be captured, and utilized for building a richer task model. An approach similar in spirit, but more limited, is described in [21] where seen text snippets are associated with files opened at the same time.

Our work builds on the same idea of long-term screen monitoring, but we utilize a semi-supervised machine learning approach which learns the user intent in real time based on screen monitoring and, if available, explicit feedback. Unlike in previous works, our proposed user intent model is interactive and is employed to recommend different types of entities (i.e., people, applications, documents and topics) that best match the user's current intention.

3 SYSTEM DESIGN

Based on the current trends in information retrieval and envisioned scenarios, we hypothesized that an effective interface for proactive recommendation support in everyday digital tasks should support entity-centric recommendations but also be able to handle fragmented data, enable efficient use of relevance feedback, and promote the generation of insights that are actionable.

3.1 Design Goals

- *Entity-centric approach.* The items recommended to support everyday digital tasks should be actionable. They should not be limited to documents but include various kinds of information entities – such as people, applications, documents, and topics – that can be used to represent the task but also to perform specific actions related to the tasks. For example, opening a document, contacting a person, or searching for resources associated to given topics. Moreover, the user interface should include hyperlinks that permit direct access to the recommended items.
- *Task-related context.* The system should be able to capture task-related contextual entities across heterogeneous applications without requiring special application-dependent customization.
- *Interactive feedback and learning.* The system should provide users the possibility to affect the recommendation through interactive relevance feedback. The use of relevance feedback should be efficient. In particular, reliance on explicit feedback should be minimized. The user interface should provide easy mechanisms to provide explicit relevance feedback when needed.

The practical implementation of the system consists of three main components: digital activity monitoring, user interface, and the user intent modeling algorithm.

3.2 Digital activity monitoring

Collecting real-life data and tasks is a prerequisite for making relevant recommendations. We aimed at a methodology that is able to unobtrusively collect all possible digital activities on a user's computer. Capturing the text read by the users on the computer screen offers a great potential to capture all important information that includes all visually communicated input and output (i.e., visual content that is generated and presented to a user on a computer screen). While screen monitoring is capable of capturing all possible information across application boundaries, it is also to capture user ongoing task-related context used to infer user potential search intents as well as can used as user implicit feedback to rank and retrieve personal documents and applications. Apart from audio, it captures all user inputs and presentation of content that occurs on the computer screen, and is thus closely aligned with the user's actual experience of the human-computer interaction.

The digital activity monitoring system is comprised of four components: Screen Monitoring (SM), Optical Character Recognition (OCR) system, Entity Extraction (EE) system, and Operating System (OS) logger.

- *SM* captures screenshots of active windows at 2-second intervals or alternatively captures the text read by the users on the screen. SM is developed into two versions: a Mac OS version and an MS Windows version. We utilized the Core Graphics framework to implement the Mac OS version, and the Desktop App UI to implement the MS Windows version. Both

perform an identical function which saves the screenshots of active windows as images. SM only captures screenshots that indicate information changes on the computer screen. Any keystrokes or changes in a computer mouse's behavior (i.e., clicks, scrolls, zoom in/out) cause SM to activate and wait for 2 seconds until no further inputs from the user, and commence capturing a screenshot of the active window. In this way we could avoid taking screenshots when the computer was idle, thus conserving CPU usage.
- *OCR system* detects and extracts text from the screenshots. We utilized Tesseract 4.0 [1] which is a commonly used and very accurate OCR implementation.
- *EE system* detects and extracts available entities from the OCR-processed screenshots. We utilized the IBM Bluemix Natural Language Understanding API [2] to extract two types of entity which were people's names, and keywords.
- *OS logger* collects information associated to the screenshots recorded such as, names of active applications, titles of active windows, available URLs of web pages or available file paths of documents that are stored on the computer. In addition, OS logger also collects timestamps of when the screenshots are captured.

All OCR-processed screenshots, extracted entities, and collected OS information were encrypted and stored as log files on the laptops for further access in the later phase. The digital activity monitoring system had a pause button which allowed participants to temporarily pause the monitoring when they did not want to share some of their private activities.

3.3 User interface and interaction

The system's user interface (UI) is illustrated in Figure 2. It implements three specific features: 1) showing entities being recommended by the system, 2) allowing selection of entities of interest by the user (explicit feedback), and 3) allowing direct action on entities when relevant. In the following, we describe how each of these features were implemented in our experimental setup.

3.3.1 Showing entities being recommended by the system. Recommended entities are displayed within four rows of five items, one row per entity type, i.e., people, applications, documents and topics (as keywords). People are identified by their name under a photo-based icon when available, and a standard anonymous silhouette when not. Applications are identified by their name under a standard icon or logo of the application or service. Documents are identified by their name under an icon based on a preview of their content, with a small icon of the application used to read or edit it. Finally, topics are identified as a single keyword. In each row, recommended entities are ranked horizontally from left to right. Since the main purpose is to show a small variety of the most relevant entities, the ranking is not visually emphasized. As users perform their tasks, the system progressively updates the recommendations. These changes are reflected on the UI as entities eventually shift places and new entities replace old ones in each row. In the prototyping phase, since entities are displayed on an orthogonal grid, some users tried to derive meaning from the vertical alignment of entities across rows. To prevent that, the grouping of recommended entities by type in each row has been emphasized with a grey rectangle that acts as a container.

3.3.2 Allowing selection of entities of interest by the user (explicit feedback). When the user is interested in a specific entity among the recommendations, she must be able to express her interest in a way that informs the system so that recommendations update accordingly. To that end, every recommended entity displayed on the UI can be selected with a click. As a result, the selected entity, or entity of interest, appears in the area at the top and the overall recommendations (in every row)

[1]https://github.com/tesseract-ocr/tesseract/wiki
[2]https://www.ibm.com/watson/services/natural-language-understanding/

Fig. 2. Two states of the system's user interface. Recommended entities are displayed within four rows, here with five items each: people, applications, documents and topics. The user can select entities of interest by clicking on them, which updates the recommendations. Example: In (a), the user sees entities related to her current work. She notices figures she has made for one of her papers (a1). She clicks on "Illustrator" (an application for editing vector graphics) (a2), then on the topic "diagram" (a3). (b) As a result, the entities of interest are displayed in the top area (b1) and the system updates the recommendations accordingly with the user's selection. In the documents row, she selects an illustration (b2) that she will modify for use in her new paper.

are updated, taking the selection into account (i.e., a positive feedback on the selected entity is sent to the system). More entities can then be selected and added to the entities of interest at the top of the screen, providing an explicit way to influence the recommendations. Entities of interest can be

Fig. 3. An example scenario: A notification (1) reminds Alice she should book flights and accommodation for a conference in Hong Kong (2). Glancing at the recommended entities (3), she notices Bob who will be attending the same event. She texts him to know where he will be staying (4). After bookings have been made, she notices in the recommendations her friend Charlie who went to Hong Kong last year and sent her an email about it. Selecting the entities “Charlie” and “Mail” (5a), she immediately recovers the email among the updated recommendations (5b). Its content gives her ideas of what to do there (6).

removed from the selection by clicking the cross that appears at their upper right corner when the mouse cursor hovers their icon. Removal of an entity of interest from the selection sends a neutral feedback on the selected entity to the system, which updates the recommendations accordingly. The whole selection of entities of interests can be reset by clicking the "Clear selection" button on the right.

3.3.3 Allowing direct action on entities when relevant. An important feature of the system is to make the recommendations actionable. While work on translating recommended persons and keywords into potential actions is ongoing, the present version simply allows to directly open recommended applications and documents. Figure 3 illustrates the user interface through an example scenario.

3.4 Learning and recommendation

The learning method receives the logged data from the digital activity monitoring (Section 3.2) and prepares recommendations to be shown on the user interface (Section 3.3). The user then may provide explicit feedback on the recommended entities, which helps the method to update the recommendations. In the following, we first discuss the modelling challenges in learning the user intent and recommendation of entities and then explain the details of our proposed solution.

3.4.1 Overview. We define context as a vector that represents all digital activities on user's computer at each time step. The set of digital activities in our setting includes OCR-processed screenshot, recorded OS information, and extracted entities from EE systems. Giving this definition for context, the user intent is modelled by a function that maps all the entities and all the previous (and also unseen future) context vectors to continuous relevance values. Learning this function is challenging since the number of entities and potential context vectors is huge (here tens of thousands) while the learning signal from the user is rarely explicit. Any method attempting to learn this mapping based on the explicit feedback alone would have difficulty, since there are limits on how low in sample size statistical methods can go [14]. To amend the limited feedback, we make the assumption that the recent contexts (i.e., latest user's digital activities) are relevant to the user's current intention. Even with this additional learning signal, the learning method still has to overcome the inherent noise in the logged context vectors; whether it is the noise in the OCR system or false entity detections by the EE systems. Furthermore, the model needs to solve the high-dimensional learning task in real time for interactive use.

To overcome these challenges we propose a semi-supervised intent learning method that uses the logged context vectors to build a lower-dimensional representation of user intent. Learning the intent function in this dimension helps to reduce the noise in the data and overcome the problem of limited feedback. We connect the lower-dimensional intent on entities to contexts by making an intuitive linearity assumption between relevance of entities and contexts, again by exploiting the structure of the logged data. Finally, we define appropriate priors on parameters, and likelihood functions (for different learning signals) to learn the user intent given recent contexts and user's explicit feedback. The learned intent is then used for entity recommendation. The mathematical details are provided in the following subsections.

3.4.2 Exploiting logged contexts for intent representation on entities. We consider the unigram model and store the logged contexts in the matrix $X \in \mathbb{R}^{|E| \times |C|}$, where the element (i, j) describes the tf-idf weighting of entity i in context j, E and C are the sets of entities and observed contexts, and $|.|$ denotes the set size. As mentioned, E and C are huge and X is noisy; still the data in X contains valuable information about how the different entities and contexts are correlated. We are interested to find a representation for entities, such that co-ocuring entities get similar representations and at the same time reduce the dimension. To this end, we use truncated singular value decomposition (truncated SVD) to get the rank K approximation of X as $X_K = U_K D_K V_K^\top$. We then use $W_K = V_K D_K^{-1}$ to project the entities into a latent space. This dimensionality reduction in context space is justified as the logged contexts naturally contain redundant context vectors. The user model is defined as a linear model in this latent space,

$$r^E = X W_K \theta, \tag{1}$$

where $r^E \in \mathbb{R}^{|E|}$ is the vector containing relevances of all entities (we use r_i^E to refer to the i^{th} element) and θ is the K-dimensional latent user intent which will be learned in 3.4.4.

3.4.3 Connecting entities and contexts. The user intent θ, as introduced in Equation 1, only maps the entities to their relevance values. We follow the keywords-documents connection idea in [11] to connect the intent to relevance of contexts by making the assumption that the relevance of a context is a weighted sum of the relevance of entities that have appeared in it, in other words

$$r_j^C = \sum_{i=1}^{|E|} p_{(i|j)} r_i^E, \tag{2}$$

where r_j^C refers to the relevance of the j^{th} context (with some abuse of notation), and $p_{(i|j)}$ is the likelihood of the i^{th} entity being present in the j^{th} context. This likelihood is not available but it can be approximated based on the logged contexts (i.e., X). We normalize the columns of X so that elements of each context vector sum up to one and denote the resulting matrix as $\hat{X}$. Using this approximation and writing Equation 2 in a vector format gives $r^C = \hat{X}^\top r^E$. Finally, by using Equation 1 we can directly connect the user intent to contexts

$$r^C = \hat{X}^\top X W_K \theta. \tag{3}$$

3.4.4 Learning the intent. In the online phase of the study, the user can provide explicit feedback, through the user interface, to the recommended entities. As mentioned, we additionally make the assumption that the recent contexts, which the user has worked on after all, are relevant to the user's current intention. To incorporate these types of learning signals and for the modeling convenience, we assume that the relevance is a sample from a Gaussian distribution with mean value r_i^E (if it is on the i^{th} entity) or r_j^C (if it is on the j^{th} context) as defined in Equations 1 and 3. The noise of these distributions should be different for entities and contexts, since the feedback on them is very different.

These feedback likelihoods are connected through the shared user intent θ. By assuming a Multivariate Gaussian prior on θ, we can complete the Bayesian inference loop and compute the posterior of θ after receiving explicit feedback and recent contexts. The posterior has a closed form solution and its mean is used to rank all entities (of different types: people, keywords, and applications) and contexts (with their corresponding linked documents) to be recommended to the user. Details of the posterior inference are provided in the appendix.

3.4.5 Computational complexity. The main computational bottlenecks of the system are the SVD calculation and the projection of all the logged contexts and entities to the latent space. These computations can be done offline, already before the beginning of the online phase of user interaction[3]. We used the Gensim Python library [41] for fast and memory-efficient computation

[3]SVD can also be updated incrementally in real-time after receiving each context. Considering the short duration of the online phase of the experiment, we decided to build the latent space only once before the start of the experiment.

	Baseline (B)	*Control (C)*	*Experimental (E)*
Entities extracted from screen data	X	X	X
Recommendation of entities using screen data	-	X	X
Recommended entities visualized	-	-	X
Explicit user feedback on recommended entities	-	-	X

Table 2. Configurations of each compared condition. Experimental design consisted of a baseline condition, a control condition, and an experimental condition. The features are additive, i.e. the features of the baseline condition are included in control and experimental conditions. The features of control condition are included in experimental condition.

of truncated SVD. This offline computation took only a few minutes for all the participants of the study.

In the online phase, the computational cost comes from the posterior calculation which is cubic on the dimension of the latent space $O(K^3)$ (see the appendix). For K in order of hundreds the computation is immediate (we used $K = 100$ in the experiments). The source code along with a simple textual user interface is available at http://to-be-included-upon-publication/.

4 EVALUATION

The purpose of the experiment was to evaluate the quality of the recommended information, usefulness and influence of the system with respect to the tasks, and users' subjective experience with the recommender system.

4.1 User Experiment Design

The study followed a within-subject design with two system configurations:

- *Experimental (E)*: condition with the recommender system visible for the user. The input to the system is the content of the screen and explicit input via the recommendation user interface.
- *Control (C)*: condition with the recommender system running in the background but not visible for the user. The input to the system is the content of the screen.

In both conditions participants could use any application running on their laptop as they would do normally, with the only difference being the availability of interaction in the experimental condition. The two conditions were counterbalanced by changing the order in which the participants were subjected to each condition.

4.2 Data Analysis Conditions

The experimental and control conditions were designed to account for the added value of the recommender method (control condition) and the interactive recommendation system (experimental) over the information on the screen (baseline).

The *baseline (B)* condition was constructed to quantify the information appearing on participants own screens. This baseline determines the information that the user has already accessed based on the applications available on the their personal computer.

The added value of the control condition is dependent on the recommendation method and added value of the experimental condition is dependent on the recommendation method deployed as a part of the interactive system. In both conditions, the added value of the recommended information was quantified over the information that users were able to find with their own tools on their personal computers.

The control condition allowed to quantify the *relevance* of recommended information resulting from the recommendation *method* without incorporating the method into the interactive system. The experimental condition allowed to quantify the *relevance and influence* of the full system. The baselines were computed separately for each participant and condition that they were in. The features of the conditions are shown in Table 2.

4.3 Research Questions

We defined the following research questions to understand the differences between the experimental and control conditions, and what appears in participants own screens:

RQ1 Does the interactive entity recommendation provide relevant information beyond what the user can find with the present tools?

RQ2 Does the interactive entity recommendation *influence* the user's information behavior?

4.4 Participants

We recruited participants by broadcasting a recruiting email message to relevant mailing lists at our university. We provided a questionnaire in the recruiting message to collect information about the participants' background, the amount of activities that they do, and the amount of time they spent on the laptop in the past several weeks. Only respondents who used the laptop as the main device for performing their everyday digital activities were considered eligible for the study. Having high educational background was another eligibility criterion, as we assumed that people satisfying this criterion would more likely use their laptops for everyday digital activities. Overall, there were fourteen respondents who were eligible to participate in the study. One participant quit after three days due to a private reason, leaving the final number of participants to thirteen. Of these, six participants had bachelor's degrees, and seven participants had master's degrees. There were five males and eight females with an average age of 25 years (std = 5). In return for their efforts, participants were compensated with 150 euros (before tax).

Upon joining the experiment, participants were informed of our privacy guidelines, and told that the data will be encrypted and stored on the secured server at the university, and only be used for the research purposes. The research followed ethical guidelines of our University. The consent form was obtained from the participants regarding the procedure of the study, data management, and data usage policy. Participants were informed that they were allowed to withdraw from the experiment at any time, and all of their data would be removed from the server.

4.5 Tasks

Discovery of the user intent and retrieval of relevant entities concerning a user's tasks is challenging in this present study due to the real-life sparse data coming from the digital activity monitoring. Some real-life tasks can be plain and quick, and do not require accessing multiple documents, for instance, reading a book. We are interested in complex digital tasks which can be viewed as concrete sets of digital activities that share a common topical context, and involve information access to a large collection of files spanning across a variety of applications, such as word processing documents, web pages, emails, instant messages, other different types of files, and folders locally stored on the laptop.

For example, a complex digital task can be a wider ongoing project, thesis work, or software implementation that are related to many files, documents read and modified by the users through many applications, or digital communication with other people. Alternatively, a task can be composed of leisure activities, such as making travel plans that include collecting vacation ideas from different websites, checking maps, organizing, accessing related notes, and making flight itinerary.

4.6 Procedure

The experiment consists of two phases: 1) two-week digital activity monitoring, and 2) a controlled lab study. These are illustrated in Figure 4 and described in more detail in the following subsections.

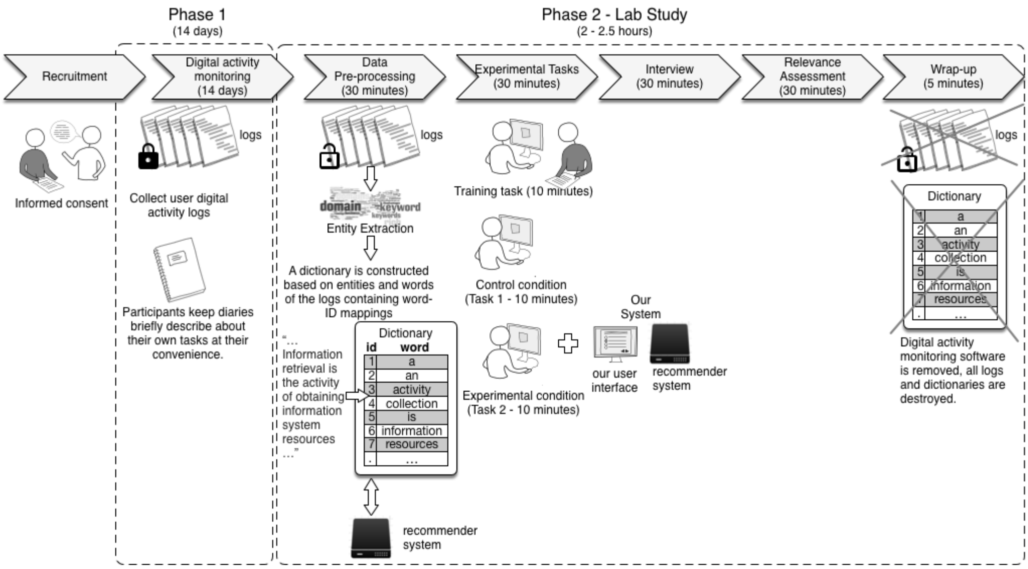


Fig. 4. Experimental procedure consisted of two phases: 1) Two-week digital activity monitoring in which logging software was installed on the participants' laptops to continuously collect digital activity logs which were encrypted and stored on the laptops; 2) Lab study included data processing, experimental tasks selected and performed by the participants, interviews to collect direct comments and impression from the participants, relevance assessment on the recommended entities, and lastly wrapping up the experiment by removing installed software, screenshots, and plain-text logs.

4.6.1 Phase One: Two-week digital activity monitoring. In the digital activity monitoring phase the logging software was installed on participants' laptops, and set to run continuously in the background thread for 14 days. Participants were advised to use their laptops as usual and avoid to pause the software unless it was necessary during this period.

Participants were asked to keep a diary describing their daily digital activities. We provided a mini booklet with a diary template including four fields: a brief statement about a digital task, related keywords describing the task, people that are involved in the task, and estimated duration of the task in a day. Participants were asked to write the diary using a pen and the provided booklet whenever they felt comfortable during a day to avoid interference with their tasks. Several example digital tasks were demonstrated to ensure that participants understood the requirements of the diary. We asked participants to focus on macro tasks that were composed of many different

Task Type	*Examples*	*C*	*E*
Data processing	Processing, analyzing, plotting, and testing data using spreadsheets, statistics software, text editors.	1	1
Thesis work	Materials gathering, thesis writing using various applications, such as file storage, repository, visualization tools, word processing, and a variety of websites.	3	2
Literature review	Reading articles, writing reviews pertaining to specific topic using pdf readers, web browsers, and word processing.	1	1
Studying	Searching information pertaining to exercises, writing a report, language studying by accessing abundant resources from online tutorials and using many tools, such as word processing, pdf reader, file explorer, etc.	3	1
Personal information management	Reviewing notes, managing files and emails pertaining to a specific topic using note-taking tools, PowerPoint, file explorer.	1	1
Programming	Project implementation, Web development, figure sketching using matlab, other integrated development environments, and various web frameworks.	2	1
Course preparation	Material gathering, and lecture slides preparation using web browsers, powerpoint, word processing, pdf readers.	1	1
Travel planning	Hotel booking, bus/flight/train ticket reservations, searches using map interfaces.	0	2
Social life	Managing group work, arranging shifts at work, communicating with others via instant messaging, email, online timetables, word processing.	1	3

Table 3. Task types, their associated examples, and the number of tasks that were selected and performed in the lab study for both conditions. C = Control, E = Experimental.

activities. For example, a digital task could be an ongoing project, or daily social networking. To ensure that the diary was appropriately composed, the participants received a weekly reminder about filling it in via an SMS.

The digital activity monitoring phase resulted in a database of 7,466 (SD=4,463) screenshots per participant. The OCR process and entity extraction resulted in an average of 1,204 (SD = 555) documents, 16,341 (SD = 7,141) keywords, 1,804 (SD = 757) people, 108 (SD = 45) applications, and 9,151 (SD = 3,351) non-entity terms.

4.6.2 Phase Two: Lab study. After 14 days, participants were invited back to our lab for phase two of the experiment.

Data pre-processing and privacy preserving: Before proceeding to the experimental tasks, the logged data was decrypted and the existing named entities inside the OCR-processed documents were recognized and annotated using the EE system. The OCR-processed documents were tokenized, split on white-space, and followed by lowercasing each word to build a corpus. We utilized Gensim library [4] to construct a dictionary which assigned unique integers (word IDs) to all the entities and words appearing in the corpus. Tokenized OCR-processed documents were converted into vectors which means each document were represented by a set of IDs using mappings between words and their IDs in the constructed dictionary (an example of a dictionary is illustrated in Figure 4). While data pre-processing was necessary for building user intent model, it was also to support preserving user privacy. Ensuring the total user privacy in this study was paramount in ensuring the participants' cooperation. Any interaction logs occurred during the experiment and assessment information collected for the evaluation were anonymized and contained no identifiable information. All the plain-text words and entities existing in the logs and assessment were converted to IDs which were unintelligible and irreversible without a correct dictionary. All plain-text logs, OCR-processed documents, screenshots, and the dictionary were not archived and destroyed upon the participants completing the experiment.

Experimental tasks selection: While waiting for the logged data to be pre-processed, we asked participants to review their diaries and select two tasks that they performed during the monitoring phase. In particular, we asked participants to pick the two tasks that they felt were similar in category, on the same level of complexity, and comparable in duration. Participants were requested to write down the description about two selected tasks in a note. After that, the experimenter randomly assigned the two tasks to the two experimental conditions. In order to counteract fatigue and other carryover effects, we counterbalanced the order in which the participants were subjected to each experimental conditions. Table 3 presents the tasks that were selected by the participants for the lab study for both conditions.

After selecting the tasks, the participants were briefed about the procedure of the experiment: they were asked to resume a task on their own laptop, engage in a short interview, and assess the relevance of the entities. Prior to starting the experiment, the experimenter set up the participant's laptop to connect to a secondary display which was an integrated 22" monitor of the SMI RED eye tracker[5]. The second screen was turned on and the eye tracker calibrated prior to the experimental condition. The recommender system's UI was set to run on the secondary monitor, and the participants performed their tasks on their own laptop. This experimental setup was designed in such a way that it was easy for the participants to observe that potentially relevant information has been retrieved, while at the same time it was easy to ignore the recommendations if the participants did not need support from the recommender system. Participants were told that they had freedom to use or ignore the secondary monitor according to their needs. In addition, OBS studio screen recorder[6] was installed on the participant's laptop to record screens of both monitors. The lab setup is shown in Figure 5.

The participants then conducted two tasks, one using each system condition, preceded by a training task.

Training task: The purpose of the training session was allowing the participants to familiarize with the recommendation system. We trained participants on how to operate the user interface. Participants were allowed as much time as needed to get familiar with the system. Training sessions typically lasted around five minutes.

Main tasks (10+10 minutes): After the training phase, the participant executed the main tasks; one with the experimental condition and another one with the control condition. After each task, we went through the video recordings of the two screens (external screen only in the experimental condition), and asked the participants to explain what they were doing during the task. For the

[4]https://radimrehurek.com/gensim
[5]https://imotions.com/smi-red/
[6]https://obsproject.com/

experimental condition, we put special focus on understanding the participants' intents when interacting with the system.

Interview: After each task, we conducted a semi-structured interview to capture participants' direct comments and impressions of the experience when using the system. In particular, interviews conducted after the second task included questions on the experiences in the two different conditions.

Relevance Assessment: Finally, we collected relevance assessments on the information that had been presented during the task. We asked the participants to assess the relevance of three sets of entities: a random sample chosen from all entities appeared on the laptop's screen including four aforementioned entity types, a random sample chosen from the same number of entities appeared on the laptop's screen from top recommendations of the user model, and top 5 recommendations of the user model (the entities that would be selected for the UI). These entities were extracted at every recommendation point which occurred at 10-second intervals. The rationale of using random sampling is to diminish user effort and fatigue from manual assessment on a large number of entities. At the end of the task, entities extracted at all recommendation points were merged, sorted alphabetically, and presented in spreadsheet file format for user convenience in providing relevance assessments. The participants rated the entities on a scale from 0 to 3 (0: not relevant, 1: low relevance, 2: medium relevance, 3: high relevance).

4.7 Post-Experiment Task Analysis

To understand to what extent the tasks in the two experimental conditions were comparable we analyzed the number of entities consumed in each condition by the various participants. The average number of screens captured and recorded was 47 (SD = 13) in the control condition, and 43 (SD = 13) in the experimental condition. On average, the number of documents was 11 (SD = 6) and of applications was 5 (SD = 3) in the control condition, and in the experimental condition participants opened 10 (SD = 6) documents and 6 (SD = 3) applications. There was an average of 411 (SD = 216) keyword entities and 34 (SD = 23) people entities that occurred during the task in control condition, whereas an average of 465 (SD = 36) keyword entities and 40 (SD = 36) people entities were found on the screens captured from the primary monitor during the task in experimental condition. No statistically significant differences were found in the number of screenshots and entities across participants between the two tasks. The results reflect that the two selected tasks were somewhat comparable.

4.8 Measures

Evaluation of proactive systems is difficult and usually done by comparing the recommendations of the proactive system with the recommendation that would have been provided if an explicit query was given to a search system [47]. This study evaluates the proposed proactive system in real-world "in the wild" digital activities which can be broader than a search task. A set of both objective and subjective measures was defined to operationalize the *relevance* and *influence* of information in different experimental conditions (see Table 4).

Recommendation relevance (RQ1)

Recommendation relevance was measured as count, precision, and the cumulative gain of novel entities that appeared during the session, either on screen (baseline) or in predictions of the user model (in control and experimental conditions). The novel items are defined as the unique entities at any point during the task, after excluding all the entities that had appeared previous to that point in the screen of the user. The rationale of this strict measure is that it characterizes the added value at the entities compared to what appeared on participants' own screens. We compared the three conditions (baseline, control, and experimental) by counting the number of entities that

Fig. 5. The interactive setup. Participants used their own laptops to perform the tasks. An external monitor was set up to connect to the laptops showing the recommender system's UI. SMI eye tracking device was installed and mounted onto the external monitor to track participants' eye gaze behavior during the tasks. In the figure, a participant continues a writing task for a research paper while the recommender system continuously suggests relevant references to the manuscript.

	Measure	Results
Relevance	Number of novel-relevant entities in equal-sized random samples (in B, C, and E).	Figure 6
	Average precision and CG of top five recommended entities (in C and E).	Table **??**
	Semi-structured interview about the quality of entities (in E).	Section 5.5
Influence	Number of novel-relevant entities occurring in the screen subsequently after they were recommended (in C and E).	Figure 9
	Duration of the gaze fixations on the recommendation screen (in E).	Figure 8
	The amount of feedback and clicks on the recommended entities (in E).	Table **??**
	Semi-structured interview about the influence of entities (in E).	Section 5.6

Table 4. Summary of considered measures in different experimental conditions.

are both novel and relevant (non-zero user assessments) in equal-sized random samples of each condition. The sampling was done by first counting the total number of appeared entities on screen, say N, and then selecting $K = \min(10, N)$ samples from the screen (for baseline condition) and K samples out of top N predictions of the control and experimental conditions. Additionally, we compared the average precision and cumulative gain (CG) [25] over the novel items in the top five ranked recommendations of the control and experimental conditions.

Influence of recommended information (RQ2)
The influence of the recommended information was measured via three types of measures targeting human interaction with the system. The rationale is that in order for the system to be useful for the users, it not only needs to recommend relevant information, but it must also influence the user either to increased attention, interact explicitly with the recommended information, or use the recommended information during the task.
Attention on recommended information was measured as the total duration of the gaze fixations of the participants on the recommendation screen.
Interaction with recommended information was measured as the amount of explicit feedback on the recommended entities and the number of opened recommended documents.
Utilization of recommended information. Use of the recommended information was measured by quantifying the subsequent utilization of the recommended entities. The rationale was that in order for the system to assist the user in performing the task more effectively, the system should recommend entities that are actually used by the user in the task. The usage was measured as the number of novel-relevant entities occurring in the user's screen subsequently after they were recommended. Only unique entities were counted. That is, if multiple recommendations correctly predict some entity that is utilized subsequently, it is counted only once.

Subjective experience of relevance and influence (RQ1 and RQ2).
Subjective experience of relevance and **subjective experience of influence** were investigated by using a semi-structured interview. The interview explored aspects related to the participants' impression that the system had influenced the task, including the quality of the experience of use, and the overall experience of relevance of the items displayed. Participants' answers were transcribed and underwent a thematic analysis. Transcripts were reviewed, and recurring themes were identified and organized into a codebook. The codes were then applied to the corpus of data [19].

5 RESULTS

5.1 Recommendation relevance

Figure 6 illustrates the recall of relevant items appearing on the user screen and in the user model in both control and experimental conditions during the lab study. For documents, applications, and persons the user model has been able to find more relevant items compared to the user screen in both the control and experimental conditions.

More specifically, there was a significant difference in the recall of applications for the user model (mod) and the user screen (scr) sources in both control ($M_{mod} = 0.68$, $SD_{mod} = 0.15$, $M_{scr} = 0.41$, $SD_{scr} = 0.19$, $t(12) = 3.99$, $p = 0.002$) and experimental ($M_{mod} = 0.78$, $SD_{mod} = 0.15$, $M_{scr} = 0.29$, $SD_{scr} = 0.10$, $t(12) = 7.9$, $p < 0.001$) conditions;

todo: double check the values Summing over the entity types, the experimental and control conditions retrieved significantly more entities compared to the baseline (70 more entities in E with p-value = 0.0014, 46 more in C with p-value = 0.0024). Furthermore, on average, the experimental

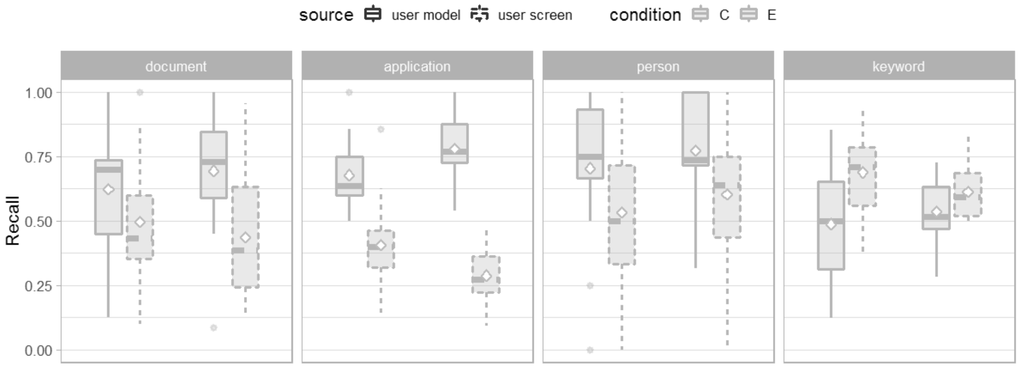


Fig. 6. Recall. Paired comparison between user screen and user model in both control and experimental conditions.

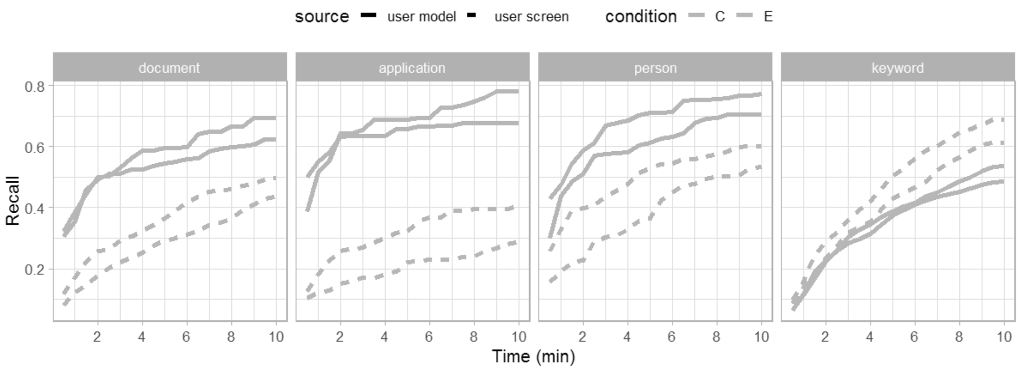


Fig. 7. Recall over time for user screen and user model in both control and experimental conditions.

condition (user model with explicit feedback) managed to predict more novel-relevant items compared to the control condition (right column). This effect is also evident by comparing the precision and CG of the top five predictions in Table **??**. The difference was not at a significant level between the control and experimental conditions[7].

5.2 Attention on recommended information

Table **??** presents the summary of user interaction with the recommendation user interface. On average, participants spent 1.3 out of 10 minutes attending to the recommendation user interface. This indicates that the recommendation user interface was used, but the participants still tended to focus more on the main task with 87% of the task duration. As expected, most of the task-related activities were performed with the participants' own laptops. A heatmap visualization of the gaze fixations is shown in Figure 8.

[7]Baseline condition is discarded in Table **??** due to not having a natural ranking measure for samples.

	Recall			P@5	nDCG@5	Influence	Feedback
	User Model	User Screen	Comparison	User Model	User Model	User Model	User Model
Applications							
Experimental	0.78 (0.15)	0.29 (0.10)	$\boldsymbol{p < 0.001}$	0.55 (0.23)	0.69 (0.01)	0.46 (0.44)	1.62 (2.40)
Control	0.68 (0.15)	0.41 (0.19)	$\boldsymbol{p = 0.002}$	0.48 (0.22)	0.68 (0.02)	0.17 (0.17)	
Comparison	$\boldsymbol{p = 0.04}$	$p = 0.10$		$p = 0.48$	$p = 0.83$	$\boldsymbol{p = 0.04}$	
Documents							
Experimental	0.69 (0.26)	0.44 (0.25)	$p = 0.09$	0.73 (0.27)	0.67 (0.04)	0.26 (0.28)	1.46 (2.03)
Control	0.62 (0.25)	0.50 (0.25)	$p = 0.36$	0.65 (0.30)	0.58 (0.03)	0.05 (0.10)	
Comparison	$p = 0.42$	$p = 0.52$		$p = 0.51$	$p = 0.30$	$\boldsymbol{p = 0.04}$	
Persons							
Experimental	0.77 (0.21)	0.60 (0.28)	$p = 0.13$	0.39 (0.34)	0.49 (0.01)	0.04 (0.06)	1.62 (2.57)
Control	0.70 (0.30)	0.53 (0.31)	$p = 0.28$	0.23 (0.21)	0.40 (0.02)	0.11 (0.28)	
Comparison	$p = 0.48$	$p = 0.41$		$p = 0.10$	$p = 0.36$	$p = 0.38$	
Keywords							
Experimental	0.54 (0.12)	0.61 (0.11)	$p = 0.25$	0.46 (0.28)	0.52 (0.05)	0.01 (0.01)	0.23 (0.60)
Control	0.49 (0.22)	0.69 (0.16)	$p = 0.07$	0.42 (0.26)	0.48 (0.07)	0.01 (0.01)	
Comparison	$p = 0.45$	$p = 0.19$		$p = 0.74$	$p = 0.62$	$p = 0.73$	

Table 5. Results: 1) Recall of relevant items appearing on the user screen and in the user model in both control and experimental conditions during the lab study. Apart from keywords, the user model was consistently able to find more relevant items than the users would find on their own. The difference was significant for the entity *application*. Additionally, on average, the user model found more relevant items in the experimental condition. The difference was significant for the entity *application*; 2) P@5; 3) nDCG; 4) Influence; 5) Average number of feedback clicks on entities.

5.3 Interaction with recommended information

During the attendance on the recommendation user interface, on average, the participants provided explicit feedback on 4.92 entities; on average once in every 16 seconds during the attendance on the recommender user interface. Most of that feedback was targeted at applications, people and documents. Participants opened, on average, 1.08 recommended documents to support them performing the task and all opened documents were assessed highly relevant with a score 3. In summary, the results indicate that the recommendation user interface was used more than 10% of the total task duration for implicitly perceiving information, explicitly providing feedback to direct the recommendation process, and to access important documents related to the task.

5.4 Utilization of recommended information

Figure 9 provides a quantitative indication of how the information provided by the recommender system was utilized in the remaining task. A paired-samples t-test indicated that the entities shown on the recommender system were effectively used to accomplish the task.

More specifically, the total number (all entity types) of recommended entities that participants used in their task was significantly higher (p-value = 0.058) in the experimental condition (in which the entities were actually shown) than in the control condition (in which the system was running in the background). This result indicates that the recommended entities were used in the task because participants saw them on the proposed system, and not by chance. In particular,

Fig. 8. Heatmap visualization of the eye gaze fixations on the recommender system's UI.

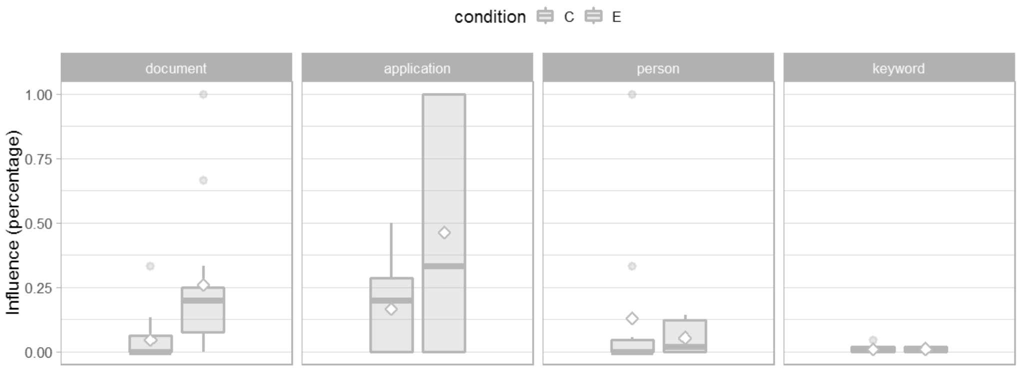


Fig. 9. Influence of recommendations on the task.

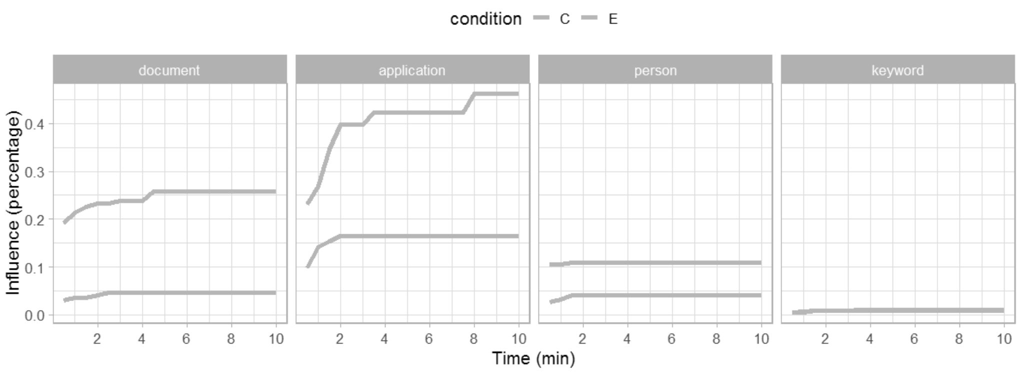


Fig. 10. Influence of recommendations on the task over time.

5.5 Subjective experience of relevance

Overall, the participants reported a positive experience with the system ($N = 9$): *"I enjoyed it... I got surprised that it recommended things that I just intended for" (P01).* All participants ($N = 13$) found the entities recommended by the system relevant:

> *They were right on the spot (P04).*
>
> *They were surprisingly relevant, various kinds of applications and various kinds of materials… I had for example this task manager, various emails, various files… they were really relevant for the subject… so it was quite good in that sense (P07).*
>
> *They were mostly very good. I think there was maybe a couple of them… they weren't relevant, but mostly it had picked up the right side (P03).*

However, some participants ($N = 2$) specified that it took a while for them to get relevant results: *"At first they didn't have much to do with what I was doing… but the more feedback I gave to the application I felt, like, relevance got a lot higher each time, so in the end they were pretty good" (P11).*

Furthermore, 10 participants mentioned that the system helped them recall specific entities useful for their task:

> *It reminded me Github. Although I would have opened it while I was working, it didn't really came to my mind… I could forget to commit my code, so it is useful (P05).*
>
> *I didn't remember that article that I got in the recommendation until I was reminded by the interface and then I was, like, 'Yay, that is actually something that's gonna help me further with the subject' (P12).*

5.6 Subjective experience of influence

The majority of participants ($N = 10$) felt that the system affected the task:

> *It made it easier to find the book, and faster (P11).*
>
> *It had a positive effect. I was thinking to remember the file name, and what I did was to just watch what the system was recommending me, and I just immediately found the name of the file. So it saved my time, otherwise I would have to find the location where I put that file. So it just simply did it for me (P01).*
>
> *Yes, it did. For example I ended up opening one of the files that was suggested, so that was something that directed quite a lot the task (P04).*

A smaller proportion of participants ($N = 3$), however, did not have the same impression that the system had affected the task, either because the task was easy and didn't require much support or because of the limited time they had for the task: *"No. I knew what I had to do so I didn't need the support" (P10)*; *"Just within 10 minutes it didn't really affect, apart from when I opened my 'Github Desktop'" (P05)*; *"10 minutes is very short for writing a thesis" (P10).*

Most participants reported that the system was useful at some point during the task ($N = 11$), typically by making it easier to access information or to recall relevant information related to the task. More specifically, six participants found the system useful in enabling faster information access:

> *It was smart enough to recognize that when I use Overleaf it would give me the book I'm using as material… and I could open the book, the link right away from there… That was a big help because if it wasn't there I would have had to go to the University's library*

website and then search the book and then go… […] and then open it… but now I could just click it right there (P11).

I was looking for a file [thesis doc] that is related to the one I have [opened in my laptop], a separate one that has comments on it, and I found it there [on the interactive system], and it was exactly what I was looking for… without… I was not writing anything, like, that I am looking for a file, it was just in my mind… but somehow it appeared (P07).

The system was found easy to use by all participants. The interface was found to be *"straightforward"*, *"simplistic"* , and *"obvious"*. Participants particularly appreciated the proactiveness of the system and how it effortlessly provided useful information:

I didn't have to do anything, it just suggests [things] for me there. So it's not yet another search-ware or anything that I need to act on it. It's just there, I don't need to do anything with it but when I look at it, it provides me relevant solutions so… easy (P07).

As interviews indicated, the majority of the participants ($N = 11$) noted a difference between the experimental conditions, and preferred working with the recommender system. Notably, four participants explained how the system was a sort of companion that helped carrying out the task with new ideas but also made it easier to stay on track:

The system gives me ideas to think out of the box… it's like talking with somebody, having a discussion that it could widen my ways of thinking how the task should be done (P02).

It's like having a notebook next to you with notes on what you should be focusing on, like a list of things you should be doing (P12).

However, some participants ($N = 4$) reported an increased effort due to the switch of attention required by the multiple screens setup: *"My attention was a bit divided because of the two screens…" (P12)*; *"I switched my attention sometime to the tool were I saw the various suggestions… it could be even not so good if I had all time switched attention" (P03).*

The majority of participants felt in control of the system ($N = 11$) because they had the feeling that the system was reacting to their digital activity. In particular, six participants pointed out that the feeling of control was mainly due to the capability of directing the system through the *feedback mechanism*:

When I checked some boxes [gave feedback] or took them off [removed the feedback], it [the system's content] changed, so, I kind of felt like it was me who affected how it worked, so I felt I was in control (P03).

Nevertheless, two participants pointed out that they would feel more in control if they could remove unuseful entities from the set of recommendations: *"I think I would feel more in control if I could actually take off the recommendations, like, the tasks [keywords] or the pages [documents] I already did, like, [those] I don't have to use anymore" (P11).* Also one participant didn't feel in control: *'No, since it's only giving me something that I can choose" (P10).*

Some participants had specific wishes or suggestions for how to improve the system. More specifically, three participants mentioned that they would prefer to have the system running on their laptops: *"If it could have been somehow embedded into the same screen… the suggestions would have been directly there where I was working" (P04).* Finally, three participants missed the capability to conduct explicit searches, at least for initializing the system: *"In the beginning it's a bit hard because you cannot set any recommendations… it would be useful to set for example some keywords right away" (P12).*

6 DISCUSSION AND CONCLUSIONS

In this paper we have investigated whether proactive entity recommendation could effectively support everyday digital tasks by automatically providing task-related entities that are relevant and useful. We introduced a user intent modeling approach where user intentions toward entities are modeled based on digital activity monitoring and explicit user input. An online system implementing this vision by using digital activity monitoring and second screen recommender system was presented. We reported results from an experiment where users were monitored for two weeks, and demonstrate the performance of entity-centric user intent modeling in improving and positively influencing real-life user tasks. We next reflect on the research questions defined earlier.

Relevance and influence of items across heterogeneous applications. Prior research on proactive agents had pointed out the need for designing techniques that were generally applicable, so the agents could be adapted quickly to different domains, corpora, and individual preferences [43]. Our study demonstrated that our entity-centric approach based on digital activity monitoring was able to effectively support users working with heterogeneous applications in their everyday digital tasks. Digital activity monitoring, in particular screen recorder built on top of applications' user interface layer, permitted to extract context across application boundaries, overcoming one of the main limitations of prior work. In our experiment, participants engaged in tasks ranging from thesis writing, data processing, and coding, to travel planning and other social tasks. Within such tasks users were involved in various activities, and their intent was often rapidly changing. Our results on relevance and influence of recommendations, together with our qualitative findings, show that the system effectively captured the rapidly evolving intent of participants and provided them with useful recommendations.

Entities help re-establishing lost connections. The entity-centric approach also helps in situations in which people have a specific item in mind but cannot recall exact pointer or name of the item. In these cases people can recall cues from their memory in order to narrow down the search, for example on whether the item is related to a particular person, or topic. Prior research has pointed out how search capabilities that retrieve information from a variety of sources, using a number of cues, in addition to keywords or folders, are critical for supporting the users in the above mentioned situations [10]. Our approach based on digital activity monitoring allowed us to capture a large set of entities in an automatic way, overcoming the limitations of more traditional approaches based on metadata [15]. Quantitative and qualitative results show that giving feedback to entities was an effective way to find the desired information, which also helped mitigate the feeling of losing control, a problem often faced when using proactive search systems [3].

Entities as interactive objects for feedback and actions. Entity selection was found to be a simple and effective way to specify one's own evolving intent when performing a task in order to retrieve the desired information. Results show that the explicit feedback was consistently used during the 10-minute experimental tasks, with an average of 3.31 selected entities. Qualitative findings suggests that this mechanism was one of the main factors contributing to the overall positive experience with the system. While the good acceptance and use of explicit feedback aligns with the idea that interactivity is an important factor to be supported in search activities that are part of wider primary tasks [26], it contrasts with other studies where feedback mechanisms were not well received [30, 57]. One possible explanation is that using entities as interactive objects could have played a main role in fostering a more active use of feedback. This intuition is confirmed by the low usage of the topic entity, which was the information item that most resembled what used in other systems.

Just-in time recommendation of useful resources that are forgotten or unknown. One challenge with complex everyday tasks is recalling useful resources that we bookmark or store for future reference. As part of their tasks people typically organize a wide range of resources that they think could become useful, so as to allow easier retrieval at a later stage. For example, as part of this, people bookmark web pages, organize emails in folders, or save documents to their hard drives. However, these useful resources are often forgotten until well after the period of their usefulness has passed [27]. With our system we remove this "out of sight out of mind" problem by automatically recommending important entities that users needed for performing their tasks, right when they need it. Results on influence of recommended information on the task show that 1.31 documents recommended in the experimental condition were subsequently used to accomplish the task. The comparison with the control condition, yielding a value of 0.31, demonstrates that in the experimental condition, on average, one document used to perform the task ($1.31 - 0.31 = 1.00$) would not have been found without the support of our system. Qualitative findings from interviews confirm the intuition given by the quantitative results, with ten participants reporting cases in which the system helped recall useful entities related to the task.

Additionally, the just-in time recommendation provided a *source of inspiration*, as indicated by interviews. This suggests that the task-related context structures extracted by the system contained insightful information that users had not noticed before.

Effortless access to information. One of the advantages of automatically predicting and retrieving potentially useful information in advance is that when the users decide to search for it, the information might be readily available without users needing to exert themselves to formulate queries. This happened several times in our study. In the interviews, participants reported cases in which the needed information was correctly predicted at the right time, based solely on the implicit interactions of the users, permitting immediate access. Other times the information could be easily retrieved after giving feedback to entities recommended by the system. Prior research points out that, although searching seems to offer obvious benefits over other methods, people still prefer to find their personal information by navigating through folders [52]. As people typically need to retrieve information items in support of a primary task, they prefer to do it in the most automatic ways, those that take less attention away from their primary tasks. Unlike searching, which requires mental effort, navigation is done mostly automatically because people are very familiar with the folder structures [6]. Our study on proactive entity recommendations, on the other hand, suggests the our method has the potential to combine the benefits of search and navigation for personal information finding. With our approach, searches where performed proactively in the background without requiring users to formulate queries, thus requiring less mental effort. Additionally, our interactive entity visualization provided visual cues that guided users toward the desired items, iteratively, by following familiar elements, similarly to what happens in folder navigation scenarios. However, our method shares the limitation of other systems for search and recommendations. Recommendations do not always come in the same order, making them less consistent than navigating through well known structures. Whether or not our method can change people's preference for navigation-based personal information access remains therefore an open question for future research.

Actionable insights. Traditional keyword recommendation may not necessarily expose affordances to the users, therefore providing them with little practical value. Entity recommendation, on the other hand, provides more actionable information items, as the item type already suggests the actions that could be done on the item. For example, seeing a recommendation of a person entity already signifies the actions that can be associated with that person, such as calling, texting, emailing, and so on. By providing hyperlinks that permitted instant access to related resources, our system made the insights provided by the entity recommendation even more actionable. Results show that, in the experimental condition, 1.08 documents used during the 10-minute experimental task were opened directly from the system's user interface. Making the entities actionable supported faster access to information, which positively affected the user experience as discussed above. However, it is interesting to note that the recommendations were useful regardless of whether people clicked on them. For example, participant P1 was engaged in a task that required to name files with a certain convention. In the interview, P1 explained that in order to retrieve that information he would have had to check out a prior project, which, in turn, would have required remembering the location of the files with the desired naming convention used in that project. Getting recommended a file with the proper naming convention, before he could even start searching, saved P1 much time, and didn't require to open the file; the mere sight of it solved the problem.

6.1 Answers to the research questions

RQ1: Does the interactive entity recommendation provide relevant *information beyond what the user can find with the present tools?* Yes, the proactive entity recommendation is able to recommend noticeably more novel (not seen before on the user's screen) relevant information than those appearing on the user screen during the actual task, with or without feedback (Figure 6) This improvement is consistently greater when the user is able to interact with the system (Table **??**). This is in line with the qualitative results from the interviews, showing that the feedback mechanism has an important role in improving the quality of recommended entities for the majority of participants.

RQ2: Does the interactive entity recommendation influence *the user's information behavior?* Yes. In the experimental condition, participants had access to recommendations that were subsequently used to perform the task (Figure 9). Interaction results (Figure 9) show that participants looked at the recommendations and used the handles provided by the system to directly access documents useful for their tasks. Qualitative feedback from interviews further confirms that the recommendations positively affected the task by reminding participants about specific activities or pieces of information that were relevant to their task. Participants generally reported an improved user experience when performing their tasks with the support of the recommender system. The system was perceived by participants as a companion that provided useful insights on how to perform the task. By reminding participants about the various entities related to the task, the system permitted to reconstruct the typical activities they performed during their task. Furthermore, the system allowed faster access to information. During their tasks participants needed to access documents of various kinds, such as pdfs, emails, code snippets, or websites. Forgetting the name or location of a file, the applications used to communicate with team members, or the subject of an important email, could make it difficult to retrieve the needed information. With our system, instead, participants could easily retrieve some important documents they needed for their tasks, either without any explicit input (for example by just opening a different relevant document), or, by selecting key entities in the system. This allowed the participants to save time with consequent improvement of the perceived experience. The benefits provided by the system came with moderate costs in terms of division of attention, as reported by some participants. All in all, the participants reported a better user experience when the proposed system was available.

6.2 Implications

The implications of the results for user modeling are striking as they open opportunities to learn user models holistically across the confines of individual applications. Our results demonstrate that it is possible to comprehensively and accurately estimate user intentions from simple input just by recording the user's screen. Moreover, our results highlight several benefits of an entity-centric approach in supporting heterogeneous tasks. Contrary to other systems which are merely based on keywords, the proposed approach permitted to extract meaning from the data and provided users with actionable insights, which, in turn, enabled easy information access, recall of the various resources related to the task, and inspiration for decision making.

6.3 Limitations and Future Work

Our work has some limitations that could be addressed in future work. While our approach based on screen content allows us to build a holistic user model from a single data source with little or no human supervision, and to provide useful entity recommendation, this method also has drawbacks. For example, active windows may include information that is always visible regardless of the particular task, such as web bookmarks. This kind of information may produce confusion to the model. Also, the model does not account for similarities between the recommended entities. For instance, a person's name can appear in different configurations (in full, abbreviated, etc.) in various applications, and thus the same people (and in general the same entities) can be recommended multiple times in the same recommendation set. Furthermore, our proposed modelling solution, in its core, uses linear models to tackle the three main challenges in the considered setup (namely, the inherent noise in the digital activity monitoring data, limited explicit interaction with the user, and real time performance necessary for interactive use). Although linear models are the preferred solution for the mentioned challenges, more complex models, such as deep neural networks, may be able to learn more complex user intents. Lastly, the data collection in the digital activity monitoring phase involved participants who turned the monitoring off for some part of their activities. Further studies of the kind of data the participants conceal on purpose would help set privacy boundaries expected by users in a more automatic way.

APPENDIX

The following likelihood functions model the three possible learning signals from the user –namely, newly generated context, explicit feedback on a document which is linked to a context, and explicit feedback on an entity:

$$
\begin{aligned}
f^{n_C} &\sim \mathrm{N}(x_n^C\theta, \sigma_{n_C}^2), \\
f_j^{e_C} &\sim \mathrm{N}(x_j^C\theta, \sigma_{e_C}^2), \\
f_i^{e_E} &\sim \mathrm{N}(x_i^E\theta, \sigma_{e_E}^2),
\end{aligned}
$$

Here the $f^{\cdot}_{\cdot}$s are the learning signals (feedback values), the $x^{\cdot}_{\cdot}$s are the projected feature vectors, and the $\sigma^2_{\cdot}$s denote the feedback noises. We distinguish entities and contexts by letters E and C, and the new observing context and user's explicit feedback by letters n and e, respectively. Following Equations 1 and 3, x_j^C is the j^{th} row of $\hat{X}^\top XW_K$, x_i^E is the i^{th} row of XW_K, and $x_n^C = \hat{v}^\top XW_K$, where $\hat{v}$ is the normalized (sums up to one) feature vector of the new context (see 3.4.3). In order to complete the probabilistic model, we put a Gaussian prior distribution on θ as $\theta \sim \mathrm{N}(0, \sigma_\theta^2\mathbb{I})$. The posterior of this Bayesian regression has a closed form solution as

$$p(\theta \mid M_{n_C}, F_{n_C}, M_{e_C}, F_{e_C}, M_{e_E}, F_{e_E}) = \mathrm{N}(\mu, \Sigma), where$$
$$\Sigma^{-1} = \sigma_\theta^{-2}\mathbb{I} + \sigma_{n_C}^{-2} M_{n_C}^\top M_{n_C} + \sigma_{e_C}^{-2} M_{e_C}^\top M_{e_C} + \sigma_{e_E}^{-2} M_{e_E}^\top M_{e_E},$$
$$\mu = \Sigma(\sigma_{n_C}^{-2} M_{n_C}^\top F_{n_C} + \sigma_{e_C}^{-2} M_{e_C}^\top F_{e_C} + \sigma_{e_E}^{-2} M_{e_E}^\top F_{e_E}).$$

Here the $F_{\cdot\cdot}$s are the vectors of observed learning signals and the $M_{\cdot\cdot}$s are the design matrices (that is, feature vectors corresponding to learning signals) from different sources.

We experimented with a pilot user, before starting the experiments, to roughly tune the hyper-parameters of the model. The parameters were fixed as $\sigma_{e_E}^2 = \sigma_{e_C}^2 = 0.01$, $\sigma_{n_C}^2 = 0.05$, $\sigma_\theta^2 = 0.1$, and $K = 100$. Positive learning signal was coded as value 1 and we considered the observed contexts in approximately the last minute of the interaction as the new contexts that are relevant to the user intention.

ACKNOWLEDGMENTS

The authors would like to thank ...

REFERENCES

[1] Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. 2015. Intentstreams: smart parallel search streams for branching exploratory search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 300–305.

[2] Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. 2018. Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing & Management* 54, 6 (2018), 1182 – 1202. https://doi.org/10.1016/j.ipm.2018.04.005

[3] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. Investigating Proactive Search Support in Conversations. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 1295–1307.

[4] Paul N. Bennett, Ryen W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. 2012. Modeling the Impact of Short- and Long-term Behavior on Search Personalization. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12)*. ACM, New York, NY, USA, 185–194. https://doi.org/10.1145/2348283.2348312

[5] Ofer Bergman, Ruth Beyth-Marom, and Rafi Nachmias. 2006. The project fragmentation problem in personal information management. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 271–274.

[6] Ofer Bergman and Steve Whittaker. 2016. *The science of managing our digital stuff.* Mit Press.

[7] Jay Budzik, Kristian J. Hammond, and Larry Birnbaum. 2001. Information access in context. *Knowledge-Based Systems* 14, 1–2 (2001), 37–53. https://doi.org/10.1016/S0950-7051(00)00105-2

[8] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. 2008. Context-aware Query Suggestion by Mining Click-through and Session Data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*. ACM, New York, NY, USA, 875–883. https://doi.org/10.1145/1401890.1401995

[9] Amir Chaudhry, Jon Crowcroft, Heidi Howard, Anil Madhavapeddy, Richard Mortier, Hamed Haddadi, and Derek McAuley. 2015. Personal Data: Thinking Inside the Box. In *Proceedings of The Fifth Decennial Aarhus Conference on Critical Alternatives (AA '15)*. Aarhus University Press, 29–32. https://doi.org/10.7146/aahcc.v1i1.21312

[10] Edward Cutrell, Susan T. Dumais, and Jaime Teevan. 2006. Searching to Eliminate Personal Information Management. *Commun. ACM* 49, 1 (Jan. 2006), 58–64. https://doi.org/10.1145/1107458.1107492

[11] Pedram Daee, Joel Pyykkö, Dorota Glowacka, and Samuel Kaski. 2016. Interactive Intent Modeling from Multiple Feedback Domains. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 71–75. https://doi.org/10.1145/2856767.2856803

[12] Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S Wang, and Alex Sandy Pentland. 2014. openPDS: protecting the privacy of metadata through SafeAnswers. *PloS one* 9, 7 (2014). https://doi.org/10.1371/journal.pone.0098790

[13] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41, 6 (1990), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[14] David Donoho and Jared Tanner. 2009. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367, 1906 (2009), 4273–4293.

[15] Susan Dumais, Edward Cutrell, Raman Sarin, and Eric Horvitz. 2004. Implicit queries (IQ) for contextualized search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 594–594.

[16] Desmond Elliott and Joemon M. Jose. 2009. A Proactive Personalised Retrieval System. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*. ACM, New York, NY, USA, 1935–1938. https://doi.org/10.1145/1645953.1646269

[17] Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Samuel Kaski, Giulio Jacucci, et al. 2013. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 117–128.

[18] Jeremy Goecks and Jude Shavlik. 2000. Learning users' interests by unobtrusively observing their normal behavior. In *Proceedings of the 5th international conference on Intelligent user interfaces*. ACM, 129–132.

[19] Greg Guest. 2014. *Public health research methods*. Sage Publications.

[20] Ramanathan Guha, Vineet Gupta, Vivek Raghunathan, and Ramakrishnan Srikant. 2015. User Modeling for a Personal Assistant. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*. ACM, New York, NY, USA, 275–284. https://doi.org/10.1145/2684822.2685309

[21] Karl Gyllstrom and Craig Soules. 2008. Seeing is Retrieving: Building Information Context from What the User Sees. In *Proceedings of the 13th International Conference on Intelligent User Interfaces (IUI '08)*. ACM, New York, NY, USA, 189–198. https://doi.org/10.1145/1378773.1378798

[22] Negar Hariri, Bamshad Mobasher, and Robin Burke. 2014. Context Adaptation in Interactive Recommender Systems. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 41–48. https://doi.org/10.1145/2645710.2645753

[23] Monika Henzinger, Bay-Wei Chang, Brian Milch, and Sergey Brin. 2005. Query-free news search. *World Wide Web* 8, 2 (2005), 101–126.

[24] Kalervo Järvelin and Peter Ingwersen. 2004. Information Seeking Research Needs Extension towards Tasks and Technology. *Information Research: An International Electronic Journal* 10, 1 (2004), n1.

[25] Kalervo Järvelin and Jaana Kekäläinen. 2000. IR Evaluation Methods for Retrieving Highly Relevant Documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '00)*. ACM, New York, NY, USA, 41–48. https://doi.org/10.1145/345508.345545

[26] Kalervo Järvelin, Pertti Vakkari, Paavo Arvola, Feza Baskaya, Anni Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Sanna Kumpulainen, Miamaria Saastamoinen, Reijo Savolainen, et al. 2015. Task-based information interaction evaluation: The viewpoint of program theory. *ACM Transactions on Information Systems (TOIS)* 33, 1 (2015), 3.

[27] William Jones, Susan Dumais, and Harry Bruce. [n. d.]. Once found, what then? A study of "keeping" behaviors in the personal use of Web information. *Proceedings of the American Society for Information Science and Technology* 39, 1 ([n. d.]), 391–402. https://doi.org/10.1002/meet.1450390143 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/meet.1450390143

[28] Evangelos Kanoulas, Ben Carterette, Paul D. Clough, and Mark Sanderson. 2011. Evaluating Multi-query Sessions. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11)*. ACM, New York, NY, USA, 1053–1062. https://doi.org/10.1145/2009916.2010056

[29] Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4189–4198.

[30] Markus Koskela, Petri Luukkonen, Tuukka Ruotsalo, Mats Sjöberg, and Patrik Floréen. 2018. Proactive Information Retrieval by Capturing Search Intent from Primary Task Context. *ACM Trans. Interact. Intell. Syst.* 8, 3, Article 20 (July 2018), 25 pages. https://doi.org/10.1145/3150975

[31] Henry Lieberman. 1995. Letizia: An Agent That Assists Web Browsing. In *INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE*. 924–929.

[32] Daniel J. Liebling, Paul N. Bennett, and Ryen W. White. 2012. Anticipatory Search: Using Context to Initiate Search. In *Proc. SIGIR*. ACM, 1035–1036.

[33] Yefeng Liu, Darren Edge, and Koji Yatani. 2013. SidePoint: a peripheral knowledge panel for presentation slide authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 681–684.

[34] Avishay Livne, Vivek Gokuladas, Jaime Teevan, Susan T Dumais, and Eytan Adar. 2014. CiteSight: supporting contextual citation recommendation using differential search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 807–816.

[35] Tariq Mahmood and Francesco Ricci. 2007. Learning and Adaptivity in Interactive Recommender Systems. In *Proceedings of the Ninth International Conference on Electronic Commerce (ICEC '07)*. ACM, New York, NY, USA, 75–84. https://doi.org/10.1145/1282100.1282114

[36] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

[37] Massimo Melucci. 2012. *Contextual Search.* Now Publishers Inc., Hanover, MA, USA.

[38] Iris Miliaraki, Roi Blanco, and Mounia Lalmas. 2015. From "Selena Gomez" to "Marlon Brando": Understanding Explorative Entity Search. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 765–775. https://doi.org/10.1145/2736277.2741284

[39] Jeffrey Pound, Peter Mika, and Hugo Zaragoza. 2010. Ad-hoc Object Retrieval in the Web of Data. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 771–780. https://doi.org/10.1145/1772690.1772769

[40] Tye Rattenbury and John Canny. 2007. CAAD: An Automatic Task Support System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 687–696. https://doi.org/10.1145/1240624.1240731

[41] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.

[42] Bradley Rhodes and Thad Starner. 1996. Remembrance Agent: A continuously running automated information retrieval system. In *The Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology*. 487–495.

[43] B. J. Rhodes and P. Maes. 2000. Just-in-time Information Retrieval Agents. *IBM Syst. J.* 39, 3-4 (July 2000), 685–704. https://doi.org/10.1147/sj.393.0685

[44] Yong Rui, Thomas S Huang, Michael Ortega, and Sharad Mehrotra. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on circuits and systems for video technology* 8, 5 (1998), 644–655.

[45] Tuukka Ruotsalo, Khalil Klouche, Diogo Cabral, Salvatore Andolina, and Giulio Jacucci. 2016. Flexible entity search on surfaces. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 175–179.

[46] Gerard Salton and Chris Buckley. 1990. Improving retrieval performance by relevance feedback. *Journal of the American society for information science* 41, 4 (1990), 288–297.

[47] Procheta Sen, Debasis Ganguly, and Gareth Jones. 2018. Procrastination is the Thief of Time: Evaluating the Effectiveness of Proactive Search Systems. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval (SIGIR '18)*. ACM, New York, NY, USA, 1157–1160. https://doi.org/10.1145/3209978.3210114

[48] Chirag Shah. 2018. Information Fostering - Being Proactive with Information Seeking and Retrieval: Perspective Paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 62–71. https://doi.org/10.1145/3176349.3176389

[49] Mats Sjöberg, Hung-Han Chen, Patrik Floréen, Markus Koskela, Kai Kuikkaniemi, Tuukka Lehtiniemi, and Jaakko Peltonen. 2017. Digital Me: Controlling and Making Sense of My Digital Footprint. In *Symbiotic Interaction (Lecture Notes in Computer Science)*. Springer, 155–167. https://doi.org/10.1007/978-3-319-57753-1_14

[50] Yang Song and Qi Guo. 2016. Query-Less: Predicting Task Repetition for NextGen Proactive Search and Recommendation Engines. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 543–553. https://doi.org/10.1145/2872427.2883020

[51] David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Susan Dumais, and Bodo Billerbeck. 2012. Probabilistic Models for Personalizing Web Search. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM '12)*. ACM, New York, NY, USA, 433–442. https://doi.org/10.1145/2124295.2124348

[52] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. 2004. The Perfect Search Engine is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 415–422. https://doi.org/10.1145/985692.985745

[53] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2010. Potential for Personalization. *ACM Trans. Comput.-Hum. Interact.* 17, 1, Article 4 (April 2010), 31 pages. https://doi.org/10.1145/1721831.1721835

[54] Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017. Proactive Information Retrieval via Screen Surveillance. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '17)*. ACM, New York, NY, USA, 1313–1316. https://doi.org/10.1145/3077136.3084151

[55] Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017. Watching Inside the Screen: Digital Activity Monitoring for Task Recognition and Proactive Information Retrieval. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 109 (Sept. 2017), 23 pages. https://doi.org/10.1145/3130974

[56] Ryen W. White, Paul N. Bennett, and Susan T. Dumais. 2010. Predicting Short-term Interests Using Activity-based Search Context. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM '10)*. ACM, New York, NY, USA, 1009–1018. https://doi.org/10.1145/1871437.1871565

[57] Ryen W White and Gary Marchionini. 2007. Examining the effectiveness of real-time query expansion. *Information Processing & Management* 43, 3 (2007), 685–704.

[58] Ryen W White, Ian Ruthven, and Joemon M Jose. 2002. The use of implicit evidence for relevance feedback in web retrieval. In *European Conference on Information Retrieval*. Springer, 93–109.

[59] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications* 55 (2015), 170 – 180. https://doi.org/10.1016/j.jnca.2015.05.010

[60] Qian Zhao, Paul N Bennett, Adam Fourney, Anne Loomis Thompson, Shane Williams, Adam D Troy, and Susan T Dumais. 2018. Calendar-aware proactive email recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 655–664.

[61] Zack Zhu, Ulf Blanke, Alberto Calatroni, and Gerhard Tröster. 2013. Human Activity Recognition Using Social Media Data. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia (MUM '13)*. ACM, New York, NY, USA, Article 21, 10 pages. https://doi.org/10.1145/2541831.2541852

Received February 2019

Publication VIII

Khalil Klouche, Tuukka Ruotsalo, and Giulio Jacucci. From Hyperlinks to Hypercues: Entity-Based Affordances for Fluid Information Exploration. In *Proceedings of the 2018 Conference on Designing Interactive Systems*, (DIS '18). ACM, New York, NY, USA, 1141-1152, 2018. DOI: 10.1145/3064663.3064728, June 2018.

From Hyperlinks to Hypercues : Entity-Based Affordances for Fluid Information Exploration

Khalil Klouche[1,2], **Tuukka Ruotsalo**[1], **Giulio Jacucci**[1]
[1]Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Finland
[2]Aalto University, School of Arts, Design and Architecture, Department of Media, Espoo, Finland
first.last@helsinki.fi

ABSTRACT

We introduce the concept of hypercue, a complement to hyperlinks in the form of an interactive representation of real-world entities (e.g., persons, places, concepts) providing personalized access points to information. Hypercues create opportunities to flexibly discover, store, and share information; organize one's thoughts; and gain insights from the data.

We explore the design space of interaction techniques supporting entity-based information exploration by reviewing recent examples of such work. We reflect on these through the lens of eight essential features of exploratory search systems to devise generalizable design principles. Our main contribution is a design template describing the hypercue, which consists of a minimal set of affordances that ensure all important features for supporting exploratory search can be addressed while leaving enough design space to facilitate integration with a variety of systems. Finally, we describe the rationale behind the design template and discuss its implications.

ACM Classification Keywords

H.5.m Information Interfaces and Presentation (e.g. HCI): Miscellaneous; H.5.2 User Interfaces: Prototyping; H.3.3 Information Search and Retrieval: Search Process

Author Keywords

entity search, interaction design, information exploration, fluid interaction

DIS '18, June 9–13, 2018, Hong Kong

DOI: https://doi.org/10.1145/3196709.3196775

INTRODUCTION

The amount of available information keeps growing exponentially, and the access points provided by query-and-response search engines – the ten blue links – are too narrow to offer a sensible overview of available material related to a given query. We need options for broader and more personalized access to information, and help making sense of it. Thankfully, new technologies in information retrieval have created opportunities to address these problems and rethink online media access and structuring. As entity search and recommendation become a reality and as recommender algorithms become pervasive, users' information trails rely less on documents linked explicitly by content creators.

Current work on search and recommendation technologies is mostly focused on the development of adaptive and autonomous user interfaces that ease the search process by reducing explicit user inputs at the expense of transparency and user control. An opposite approach would be to reinforce user control through the use of direct manipulation and rich user interfaces that provide users with a feeling of accomplishment and increased responsibility over the search process, as advocated by Ben Shneiderman [29]. Development of such systems is not new [1] but their reliance on filtering methods often limits them to homogeneous datasets and specific use cases (e.g., movie exploration and product finding) that limit opportunities for discovery and generalization.

However, as contemporary search and recommendation is fueled by extremely complex data structures, the same technologies create the opportunity to develop interactive systems that provide users with the ability to finely steer their progression within the information space in accordance to their immediate needs, understanding and inspiration. Related information and overview of the data can be computed on the fly to suit the very specific needs of each user at any time during the exploration, providing constant access to more detailed or more general information, new directions and branching topics. To enable such possibilities, we need visualizations and affordances that do not rely on preestablished criteria to further the exploration.

The need for supporting information exploration has been demonstrated by studies and models of information-seeking behaviors. Features of such support have been described, and most information tools (e.g., websites and browsers) al-

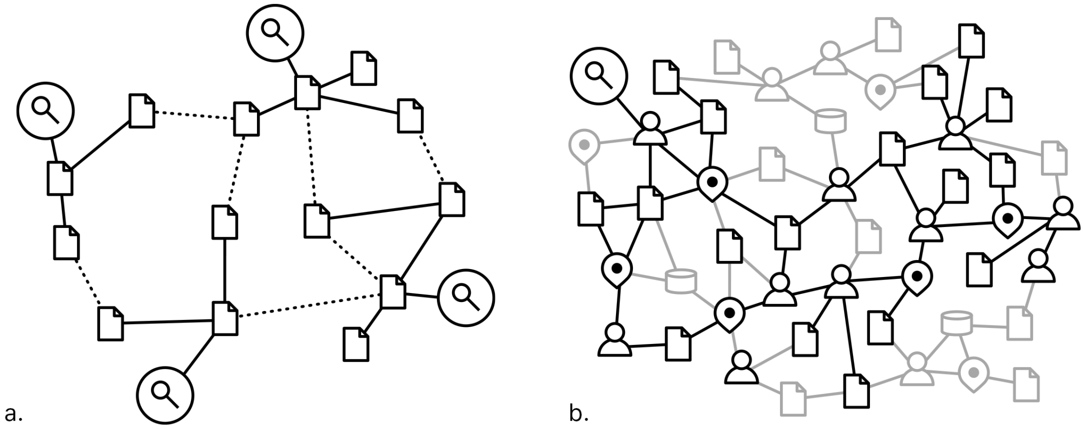


Figure 1. Information exploration is often compared to a journey through the information space. In practice, however, text-based querying and the limited overview force users to take multiple discrete steps inside the information space to make sense of it (a). In our research, we aim to utilize the underlying data structure to help users make sense of the information space and foster insights (b).

ready implement such features by providing shortcuts to quick lookup, related items or info boxes when relevant. Our research builds upon such streams of work to develop actionable solutions for implementing said features. However, instead of taking an additive design approach that would consist of bundling together multiple widgets, each addressing a specific task, we are more interested in the exploration of alternative interaction paradigms and the discovery of novel key affordances to take full advantage of the volume and complexity of the data available today. Although entity-based exploration has been studied and is already commonplace, it mostly consists of following information paths provided by successive single recommended items, like following a path of hyperlinks. Such an approach can lead to enjoyable serendipity, but it lacks the means to show the personalized results that a typed query can provide. As for typed queries, recalling search terms to formulate and reformulate queries (especially when exploring an unfamiliar information space), is not as easy as following visible items. Moreover, although information exploration is often compared to a linear journey through an information space (e.g., the information foraging theory [23] or the berrypicking model [3]), in practice, text-based querying and limited overviews force users to take multiple discrete steps inside the information space, as seen in Figure 1a, instead of utilizing the data's rich underlying structure as seen in Figure 1b. Our work suggests that a few key affordances, such as the ability to compose queries by grouping multiple entities and then refine it as new entities are discovered, provide a compelling and efficient way to explore a large information space while taking advantage of its entity-based structure.

This conceptual work is a reflection on our exploration of interaction techniques designed to support entity-based information seeking, grounded in available literature. Our main contribution is a design template describing the hypercue, an interactive representation of entities that provides personalized access points to information and serves as a complement to hyperlinks. Hypercues create opportunities to flexibly discover, store, and share information, and to gain insights from the data. The hypercue design template consists of a minimal set of affordances that ensure all important features for supporting exploratory search can be addressed while leaving enough design space to facilitate integration with a variety of systems. We describe the rationale behind the design template and discuss its implications.

BACKGROUND

A large body of work builds upon what seems to be a human propensity to consider the information space as a physical space in which we move, progress and discover. Bates' berrypicking approach to search [3] and the information foraging theory [23] describe users' information-seeking behavior through the metaphor of a physical journey. With the "information flaneur", [6] Dörk et al. go deeper into the physical metaphor by comparing information spaces to the 19th-century city in terms of growth, cultural significance, and being the place for social struggle and negotiation, and they use such parallels to envision "positive" approaches to information practice. Such models have all contributed to shift the emphasis from a mostly technical consideration of information retrieval toward human processes [14]. We see the present work as the design counterpart of such a framework, providing actionable principles toward implementation of systems supporting and based on observed human behaviors.

In this work, we utilize entity search as technical opportunities supporting fluid information exploration as a goal. Here, we clarify these two essential notions.

Fluid Information Exploration

When we discuss interactions and user interface design, fluidity is often cited as a goal [10, 25]. White and Roth [31] mention fluid interactions as an important feature of future search systems when discussing novel interaction paradigms. They link that notion to human-machine symbiosis and interactions through fluid hand gestures, citing the fantasy user interface from the movie *Minority Report* as an example of what a truly fluid interface could look like. However, that notion is not theoretically defined and is generally used while relying on the reader's intuitive understanding of the metaphor, something that flows continuously, naturally making its way around obstacles and adapting its pace to the environment.

A satisfying operational definition is proposed by Elmqvist et al. [7], who work around the difficulty of defining fluidity theoretically by focusing on the properties we can expect from fluid systems. These properties are grouped into three sets:

Fluid interactions support direct manipulation: Direct manipulation describes an interaction paradigm in which digital representations of objects behave as objects themselves [28]. Direct interaction with these objects is enabled by reducing indirections between input and output spaces. For example, the touch-sensitive layer of a touch device is confounded with its display and calibrated so that inputs are registered precisely at the display location. The paradigm relies on a preference for physical actions, and immediate visible effects allowing rapid course adaptation [13].

Fluid interactions promote flow: Flow is a mental state induced by immersion in one's activity, characterized by a loss of sense of time. The main actionable property for inducing flow relies on letting users feel in control, and employ just the right amount of skills to let them progress in their tasks at a pace that will feel neither too slow nor too fast, accommodating a person's continued and deepening enjoyment as skills grow [20].

Fluid interactions minimize the gulfs of action: The gulfs of action are a notion introduced by Donald Norman [21], who uses it to describe the gap between a user's expectation of a system and the system's actual state.

A limitation of this definition is that these properties are not complementary, as they seem contained in Shneiderman's advocacy for comprehensible, predictable and controllable user interfaces through direct manipulation [29]; however, it provides various approaches to the implementation of such a goal.

Our focus on fluidity is not exclusively motivated by performance or the sole objective of making the user more efficient in accomplishing her task. Fluidity conveys open-endedness and the consideration that the process itself, through the potential discoveries it may yield, is often just as valuable as the end product. The information flaneur [6] offers an inspiring model of information seeking centered around positive information practices, as opposed to considering information seeking the fulfilling of an information need or addressing of a deficiency. The information flaneur's implications for research and design offer design goals for fostering or enabling such experiences by considering explorability principles (e.g., orientation, visual momentum, and opportunities for serendipity) and bridging gaps between information spaces, contexts, and conceptual levels by exploiting scalable or generalizable rules and common patterns. This focus on continuity and momentum of the exploration experience complements the operational definition of fluidity.

The present work reflects our commitment to such an approach to information practices, as we strive to design interactions that do not induce specific behaviors nor limit their utility to very specific scenarios. Given a specific scenario or task, dedicated tools might undoubtedly outperform a more generalist approach applied in a monolithic system. However, we envision a future of information exploration where users' interfaces of choice are not constrained by the data but by personal preferences that could indifferently and seamlessly be used to navigate academic literature, a movie streaming catalog, the news or social media posts, allowing the transfer of any progress from one tool to another thanks to a set of standard affordances like the ones described in the present work. We work toward that goal by implementing the principles of fluidity mentioned above. We support direct manipulation by enabling direct interactions with every bit of displayed information that is relevant to the user. As a result, we design for touch-enabled displays, and present information in the form of objects that can be manipulated. We build interfaces around small sets of simple rules that are consistent across a given system, which we achieve by limiting the amount of widgets and separate views and preferring single workspaces. Through these principles, we attempt to facilitate memorization regarding operations while enabling creative strategies and behaviors leading to potentially complex results.

Entity Search

In the information search field, entities are references to real-world objects or concepts (e.g., persons, places, movies, topics, and products). Entities are linked with typed relationships. For example, "Tom Hanks (actor)" and "*Forrest Gump* (movie)" are linked via "stars in". Together, they form a graph in which entities are nodes and relationships are the edges. Such graphs are known as *knowledge graphs* and generally stored within *knowledge bases*.

In web searches, a majority of emitted queries pivot around a specific entity [24]. Knowledge bases are useful to providing additional information around an entity and recommending additional entities. In conventional web search engines, queries pointing toward an entity will usually trigger a first result that points to the information source with the most general information about the entity (typically, the corresponding Wikipedia entry), which in turn requires the system to be able to match the typed query to the corresponding entity.

Entity-Based Exploration

Google Search provides for most entity-based queries not only a relevant entry but a knowledge graph with relevant information about the entity and recommended related entities in the case of an actor, name, age, and lists of movies and costars. Miliaraki et al. [18] studied the behavior of users of Yahoo Spark, a system that recommends related entities alongside Yahoo Search results; the users take advantage of the system to engage in explorative entity search by discovering information through successive clicks on recommended entities. Such cases exemplify why entity-search is considered an ideal paradigm for exploratory search and an important topic in information retrieval and semantic web communities. A large body of recent research work addresses challenges

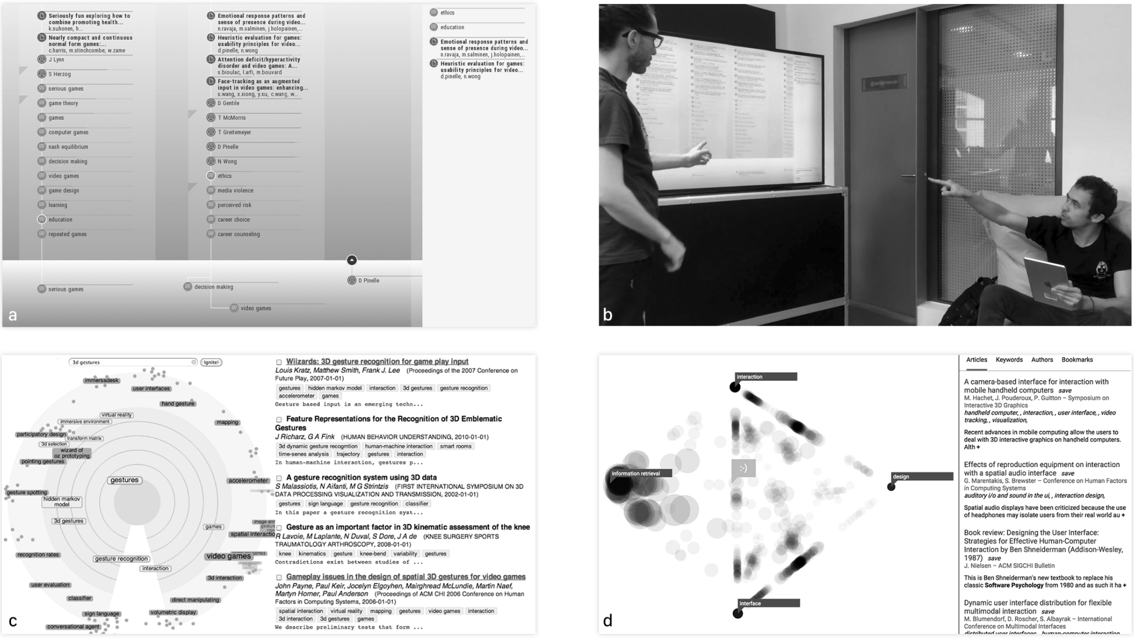


Figure 2. We have explored the design space of entity-based systems supporting information exploration through the development and evaluation of multiple prototypes: (a) ExplorationWall, (b) QueryTogether, (c) SciNet/IntentRadar, and (d) RelevanceMap.

regarding the computation necessary for entity search, such as the finding and ranking of related entities, matching entities with occurrences in free text queries and completion of entity lists based on given entity examples. However, as techniques improve, it is difficult to find research addressing interaction techniques that enable end-users to access and benefit from such rich information in a wide variety of activities, as examples usually target very specific scenarios and tasks. Several such entity-based exploration systems have been designed to support expert investigators in making sense of a corpus of documents [4, 30, 5].

As entity search becomes a reality, it creates plenty of opportunities for exploration, not only in the way relevance and ranking are computed but in how we will interact with information within the new paradigm. Once available information is meaningfully structured, the information space becomes a high-dimensional medium ready to be unfolded as explorers pull its various threads in various directions and discover content according to their needs, inspiration or chance. The goal of entity-based fluid information exploration requires substantial thinking about the way we display and interact with entities and come up with fundamental principles that are generalizable to any search contexts (e.g., academic publications, social media, movie database and personal emails).

EXPLORING THE DESIGN SPACE

We have had the opportunity to explore the design space of systems that support fluid information exploration through the design, development and evaluation of several functional prototypes. These prototypes each addressed different aspects of information exploration (e.g., facilitating query formulation, facilitating query refinement, providing insights from the data and supporting collaboration). Through user studies, we have been able to demonstrate improvement of these various aspects over conventional approaches and identify design principles that were responsible for these improvements. We have also been able to observe how these principles affected user behaviors and search strategies. The next challenge was to use this collection of observations on individual systems, to extract fundamental actionable principles that would outline a paradigm for fluid information exploration.

ExplorationWall

Figure 2a. shows ExplorationWall's interface [15]. This work initially addresses challenges in performing exploratory search tasks using touch-based devices, such as formulating queries in unknown information spaces, identifying new search directions, and going through long lists of results with low information gain. Those challenges are all made more difficult on conventional user interfaces by the lack of physical text-input or text-selection peripherals.

A single workspace allows for the simultaneous display of several parallel search streams, each consisting of a vertical organization of entities as queries and search results. Each result set consists of multiple entities of varying types (i.e., documents, persons and keywords) and each entity can in turn be used as a query or part of a query in an existing search stream or in a new one. Entities can be easily manipulated (e.g., moved, stored, or combined) and can provide content to be browsed, such as articles or insights to extend exploration(e.g., new topics or authors).

ExplorationWall has demonstrated substantial improvement over regular interfaces in exploratory search tasks [15]. Results showed a higher amount of relevant information retrieved by participants explained by a much more active behavior and measured in the search trail analysis in number of queries, as well as revisit and branching rates. In the end, participants that were given a topic to explore with which they were unfamiliar, covered much more of all available relevant information by making better use of multiple search sessions.

QueryTogether

QueryTogether has been designed to support exploratory search in a collaborative and spontaneous search setting, as shown in Figure 2b. To do so, we have adapted ExplorationWall's interface so we could use it across multiple devices (e.g., laptops, tablets, and large touch displays) and support both private sharing of information and broadcasting to all users.

The reading list has been adapted to feature saved and received entities in a scrollable list and a user panel, which shows what users are active in the system. Users are identified by the name they entered at login and their status, either *private* if they are using a private device or *public* if they act as a session moderator on a large screen available to all participants.

Sharing is performed by dragging an entity over to the chosen user. The recipient will instantly receive a new instance of the sent entity in her side panel. If the side panel is closed, a visual notification in the corner informs the user of the number of new entities received. Next to each user label, a "Message" icon allows the user to send a short message along with an entity. The reading list can be filtered with respect to a chosen collaborator and will subsequently show only entities and documents sent to or received from that user. Filtering based on one's own ID will display only entities that have been saved locally and ignore anything sent or received remotely.

SciNet/Intent Radar

SciNet/Intent Radar is a search system for information exploration that builds a user intent model to better adapt returned results to a user's needs [8, 26]. To do so, the system enables relevance feedback on a visualization of the user's intent, which is here visualized separately from the result list through the use of weighted keywords, as seen in Figure 2c. The benefits of a separate visualization are a higher cue density, a comprehensive overview of the current intent, the opportunity to provide suggestions for future intents and the practicality of being an add-on widget to familiar search interfaces. In this case, a user provides an initial query, which yields a regular result list, plus a set of the most central keywords extracted from the results, visualized on a radar (closer to the center meaning more central). This interface offers an overview of how the system perceives the query. The user can then adjust the weight distribution by sliding individual keywords closer to or farther from the center, and the result list will refresh accordingly. Another one of the system's important features is the visualization in the radar's outer ring of a variety of secondary keywords, chosen for their diversity, providing the user with insights about future search directions and a way to redirect the search incrementally without having to change the initial query. The system has been shown to improve users' task performance in complex search tasks in which conventional query-response systems failed to help users direct their search [26].

RelevanceMap

RelevanceMap [16] provides the user with an interactive map of the whole document space with respect to the positions of multiple query phrases visualized as mobile markers on a 2-D workspace, as seen in Figure 2d. The map allows for visualization of variations in information density with respect to elements of the query, even with large amounts of returned documents, which enables quick evaluation of the query and the resulting data. Browsing is performed through a pointing gesture on the map, which re-ranks the whole document collection according to the location of interest. The corresponding result list appears in a conventional layout next to the map. The quick re-ranking interaction enables exploration of the multi-dimensional data structure.

This work is another example of visualization taking advantage of direct manipulation to interact with the data for sensemaking purposes. In this case, however, manipulation of the map allows the user to determine the scope of the higher level search task by outlining the whole document space using multiple queries and visualizing the whole set (i.e., thousands of documents instead of a small selection). The exploration is then performed through a re-ranking interaction over areas of interest on the relevance map. The user can now precise or change the focus of her exploration and refresh the result list with a pointing gesture instead of reformulating the query. This ability allows the user to explore large amounts of relevant information without interruption, while still permitting changes in the query at any time, thus adapting fluidly to the user's immediate needs.

RelevanceMap showed significant support in how users perceived the information space with respect to topics of interest and in retrieval of information relevant to complex criteria [16].

FEATURES OF EXPLORATORY SEARCH SYSTEMS AS A LENS

In our exploration of the design space of systems that support information exploration, we have often taken advantage of the work of White and Roth [31] as a frame to ground our designs and explain their effects on user behavior and performance. Their work includes a list of features of exploratory search systems that exemplify essential aspects of supporting information exploration. These features have been an important source of inspiration as they bring attention to various fundamental areas of exploration.

We reflect on our work through the lens of these eight features of exploratory search systems to identify fundamental principles at work in our various systems and new opportunities for entity-based exploration.

Support Querying and Rapid Query Refinement

Search tasks are commonly addressed by inputting queries in a search system, which then yields a set of related results. However, conventional text-based queries are mostly user-defined. Relying on the user's existing knowledge to formulate satisfying search directions limits the range of incrementation in the iterative exploration process. Support for querying is commonly addressed by providing the user with ideas for new queries or additional terms. Even auto-completing text entries with popular queries has been shown to facilitate the querying phase.

When entity search is available, supporting querying and query refinement is readily achieved by augmenting conventional expression-based querying methods through entity recommendation by enabling the use of entities as queries, to provide a set of related entities as a result. A document, a person, a place or a movie can each be used as a query and can yield a variety of related new entities of various types.

ExplorationWall showed how users embrace such support for exploration over text-based querying and how that support improves the overall exploration [15].

Facets and Metadata-Based Result Filtering

Being able to navigate a large result set according to personal needs and preferences is a central requirement of fluid information exploration. That is why this ability – to narrow down such results according to a variety of criteria that are representative of what is available in the data and complementary enough to provide a meaningful choice of search directions – is an important feature to support.

Facets and metadata-based parameters are an attempt to structure information by linking documents semantically through common features (e.g., an author, a title, a date or a location). Entity search is the ideal paradigm for result filtering, given the richness and complexity of readily-linked data. From an initial query entity, a system would retrieve the most central neighboring concepts or elements and provide them as related entities to chose from. The initial result set can then be narrowed down or re-ranked with respect to the relatedness or dependency of each element to the chosen related entity.

Each result set in ExplorationWall displays facets in the form of recommended keywords related to the query. Adding such keywords to the initial query narrows down the results. For example, we have observed a user initiating a search with the query "rosetta". Results were related to the ancient translation stone and the recent European mission to land a probe on a comet. By moving the recommended keyword "probe" to the query, the result set was then focused on the space mission. Then, after the user added the recommended keyword "instruments" to the query, the result set became a catalog of Rosetta's on-board systems.

Leverage search context

A substantial part of context can already be harnessed by accessing contextual data provided by sensors (e.g., GPS signal or personal account informations). From our interaction design perspective, we are more interested in techniques enabling inference of context through users' input, either explicit or implicit.

Explicit input of context implies providing a user with the ability to critique encountered information by providing relevance feedback that informs the system of the user's intent. SciNet implements such principles and has demonstrated substantial benefits for exploratory tasks.

Implicit input of context relies on inferences made from the user's behavior. What information is being saved for later, liked, or shared is often implicitly considered relevant to a user's interest. The songs the user put together in a playlist and the movies she has watched or positively rated are primary actions with implicit implications that create an opportunity to improve any associated recommendation or retrieval process. In systems where users can freely position selected information on a workspace, such as RelevanceMap, the layout and proximity factors can be used to infer a user's intent for disambiguation and to improve the results.

Visualizations to Support Insight and Decision Making

Interactive information visualization is an important tool for sense making. Being able to encode data visually and to play with various parameters is a powerful way of discovering trends, understanding relationships, gaining insight from the data and ultimately informing decisions. Entities in knowledge graphs generally make for inspiring material regarding visualization techniques such as node-and-link diagrams and adjacency matrices.

The challenge with information visualization is that the most appropriate form of visualization is highly dependent on the task and the data. In that respect, it is crucial to be able to adapt when necessary through various techniques.

RelevanceMap uses user-driven mapping to enable multi-aspect search. Having multiple queries distributed on a surface makes it possible to consistently map a whole information space using dimensionality-reduction techniques. RelevanceMap provided users with sensible insights from the data and with the ability to explore tradeoffs with respect to multiple criteria, thus supporting decision-making [16].

Support Learning and Understanding

As it is necessary to offer some result-filtering ability for the user to take better advantage of a large set of results by narrowing down a list, it is also important to provide the user with access to more general knowledge when necessary. Support of learning and understanding implies that a user is given the means to find information that is adapted to his current level of understanding. This is typically achieved through recommendation of related material. For example, any modern browser or eBook reader provides the ability to look up the definition of a word or to link a concept with its corresponding Wikipedia entry.

More elaborate solutions consist of yielding a variety of recommended concepts or documents related to the information at hand, which provides the user with a conceptual overview of the topic with multiple options to gain knowledge directly useful to understanding the information at hand.

SciNet, in addition to visualizing the current intent through the outer rings, displays there a variety of diverse keywords, providing the user with options for future search directions, as a way to expand the information space without having to change the initial query.

As a second example, in the same way that ExplorationWall enables facets and the narrowing down of the result set through multiple queries, similar techniques can be used to expand the result set by adding new entities to the query, this time from external sources (i.e., another result set or a typed expression).

Facilitate Collaboration

Collaborative information exploration is a common strategy to tackle large information spaces through the sharing of ideas and allocation of search tasks [11]. Collaboration can take multiple forms, with settings in which collaborators either share or do not share the same space (i.e. colocated or distributed collaboration), either synchronously or asynchronously.

Methods that support collaboration when searching or interacting with information are very diverse but often share the common goal of avoiding wasteful overlaps in labor between collaborators. This problem can be tackled at the start of a search session, through mechanisms supporting division of labour by systems such as SearchTogether [19] or Cerchiamo [9], which allocate defined search areas to each collaborator. However, the same problem can also be alleviated during the search/exploration by promoting awareness of collaborators' activities, by providing a communication channel allowing participants to share their progress in the form of interesting information snippets and sources and by making their search trails visible, as CoSense [12] and CoSearch [2] do.

Entity search creates opportunities for collaborative exploration, as the information unit of reference shifts from the sole document to a variety of references to real-world objects (e.g., persons, places, concepts, and expressions – or any combination of these). We can therefore envision systems in which users have finer control over the types of information and amounts of context saved and shared.

For example, using QueryTogether, users shared various entities to facilitate task allocation and find common ground in collocated collaborative exploratory tasks. We can imagine implementing the exchange of whole workspaces and search histories as a way to share not only bits of information but whole contexts.

Histories, Workspaces, and Progress Updates

Information exploration is a sense-making activity [17]. As such, it is open-ended, potentially long-term, and changes continuously as the information need evolves [22]. The process often produces long and complex search trails with multiple branches and revisits. In this context, it is important for a user to be able to take advantage of previously encountered information and to keep track of past activity to more efficiently recognize new and interesting information.

Entity-centric information presents an opportunity to highlight updated and visited information at a finer level than documents and in overall areas of the information space. In visualizations taking advantage of the node-and-link structure, it is possible to use color to highlight visited or new areas in the knowledge graph as we implemented in later versions of RelevanceMap, where visited areas of the information space were colored in a purple shade, as is traditional for visited links.

One of ExplorationWall's central features is the endless horizontal workspace that allows users to freely move and position parallel streams (i.e., query plus corresponding result set) and to create additional space in between streams. We noticed that not only did users take advantage of such features to thematically organize multiple search directions; their natural tendency to add new elements to the right systematically resulted in naturally grown chronological search histories.

Support Task Management

As Information Exploration is potentially long-term, it is important for users to have the ability to interrupt and resume their activity and to carry it over time and across devices. This require the ability to not only save selected information but to provide future access to whole workspaces, including histories and information configuration with which a user has engaged.

An important criticism of the way we commonly search for information is the general ephemeral and secondary perceptions of the search process. Users often invest a great deal of time and effort finding information, and the process itself, including intermediate queries and results, is often lost after an endpoint has been reached. As we advocate for improved user control over the search process, often at a higher interaction cost, it is crucial to recognize its value. This improvement implies ubiquitous computing solutions to enable the saving and resuming of search sessions across time, space, and devices. We implement such features in our systems by saving workspaces to the cloud and, in the case of QueryTogether, porting the system to multiple platforms.

THE HYPERCUE TEMPLATE

A cue is a stimulus and a signal for action. We propose the notion of a hypercue as a complement to the hyperlink. A hypercue is an interactive representation of a real-world entity; it offers affordances (i.e. possibilities for action) for the user to explore, share and organize her thoughts. Systematic inspection and exploration of the design space of each feature of exploratory search systems allowed us to identify three complementary affordances that are responsible for enabling these features and that together constitute a minimal design template for implementing hypercues. The following template aims to guide the creation of future interfaces for exploration without overconstraining the design of such systems, or hindering the ability to address specific cases through the choice of a specific form of visualization. The proposed affordances can also be implemented in most existing media-handling applications (e.g., in browsers and in PDF and e-book readers). From the user's perspective, the following template provides a base set of rules and expectations to facilitate users' engagement in complex information behavior.

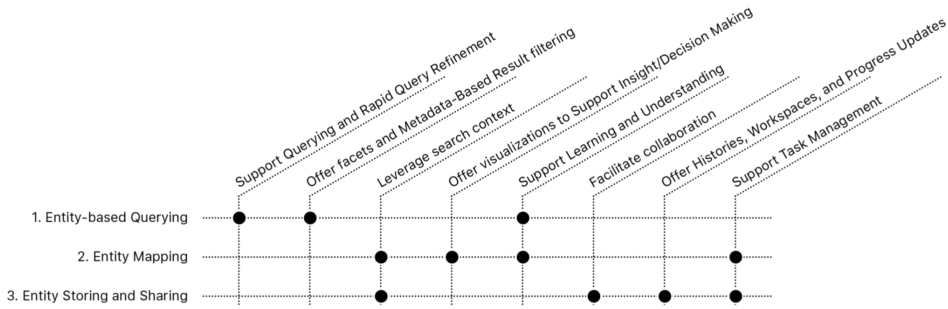


Figure 3. The three entity-based affordances for fluid information exploration together create opportunities for implementing each of the eight features of exploratory search systems that White and Roth described [31].

Entity-Based Querying

Each entity or combination of entities yields various new related entities, thus providing an overview of the respective information space.

Providing the ability to create queries through the direct manipulation of recommended entities can **support query formulation and facilitate query refinement**. The ability to add more entities to an initial query makes it possible to refine it by narrowing down or expanding the result set. Adding external entities (e.g., from somewhere else in the article or page being consulted, or from another source) results in the expansion of a query, thereby **supporting learning and understanding**. Adding an entity to the query from the result set **enables facets and metadata-based result filtering**.

Entities become resources for users to express their information interests and search intents. Sets of related or previously observed entities can be used to collect feedback from users on their current reliance, which would support advanced personalization in iterative user modeling where the exploration system presents predictions of user intent through sets of entities helping the user to discover and formulate her current intent.

Modern browsers and operating systems already implement affordances to inspect the definition of an expression, its corresponding Wikipedia entry or related search engine results. Such affordance is here generalized, using entity search to yield a crop of related entities from any selected object (e.g., an expression, article, or link).

Entity Mapping

Entities can be moved around, and users are provided with the spatial freedom to organize the entities of interest in a layout that reflects their understanding and their mental representation of the information space.

Spatial organization of thoughts is a common behavior. We draw mind maps, we make piles of documents, we organize sticky notes, and store documents within directories or under consistent tags. Sense-making is an important part of exploratory search [31], and as such it relies on users building a mental representation of the state of the world (i.e., the information space at hand) and then iteratively contrasting this representation against the real world (i.e., new information) to update it and acquire a progressively more accurate understanding of the information space [27].

Entity mapping provides support for mind mapping, which **supports learning and understanding**. It provides an implicit input channel for **leveraging the search context**. It also allows for creating **visualizations that support insight and decision-making** by enabling multi-aspect search, as well as for addressing the need for **histories, workspaces, and progress updates**.

Entity Storing and Sharing

Entities and groups of entities can be easily saved for later use and easily shared with collaborators.

Documents often serve as units of information. Users search for, bookmark, and share such documents. Such actions are not sufficient, however. The user often forgets the intent behind the bookmarking and thus loses the utility of the stored document. Some additional action is required, such as giving each bookmark a context-relevant title or organizing bookmarks within theme-specific directories. Sharing requires the use of messaging channels, as text messages are necessary to convey context and intent. Entity-centric information enables the use of variable and personalized units of information. Users can search for, store, and share references to persons, media, excerpts, and organizations. Taking advantage of affordances 1 and 2, the exchange of information involves potentially sharing – and collaborating on – whole contexts in the form of organized entities, which **facilitates collaboration**. The same principle gives access to these contexts across devices, providing flexible **support for task management** and **enabling histories, workspaces, and progress updates**. Stored or saved information also provides an implicit input channel for **leveraging the search context**.

Summary

Figure 3 shows how the hypercue template, which is based on the three proposed design principles above, creates opportunities for addressing each of the eight features of exploratory search systems.

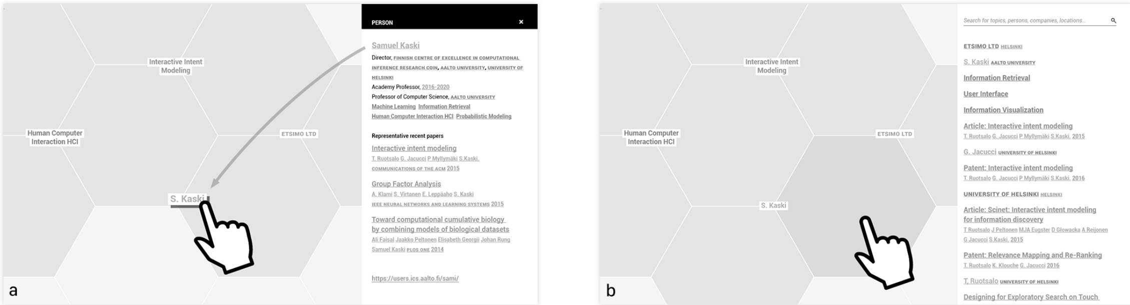


Figure 4. InnovationMap is designed for innovation exploration and discovery, and implements the three principles of the Hypercue template. (a) A user inspects the entity of a researcher in the right panel. If it is of interest to her, she drags it towards the map. (b) The right panel now displays the search results corresponding to the (blue) cell selected by the user, which has two entities as a compound query. As results are saved in the cloud, the user can resume her exploration later on a different device.

Example

To illustrate our approach to designing systems that support information exploration, and to show the hypercue's role in enabling relevant features, we describe and discuss the case of InnovationMap, a system that was designed following these principles. It is currently being implemented in partnership with the University of Helsinki to support innovation exploration and discovery.

User Interface

The workspace consists of a map, with a panel on the right, as seen on Figure 4. The map is based on a grid of hexagonal cells. The scale of the map is changeable through the use of the mouse wheel or through pinch gestures. The companion panel has a text-input field at the top, inviting typed queries, and the corresponding result set is displayed under it. Results consists of entities of various types (i.e., documents, persons, workplaces and topics). An initial text-based query yields a ranked set of mixed-type entities. Each entity can be inspected with a click or a tap, at which point the panel displays all the information about the selected entity (i.e., the full metadata) to provide a preview of the content – and access to it, if applicable, through available links such as dedicated websites, Wikipedia entries, or social media contacts. Each entity can also be dragged towards the map, where it will position itself at an intersection of the grid. The darker cells on the map are representations of the various result sets yielded from all adjacent entities. Tapping on a cell highlights it in blue, signifying that the corresponding result set is displayed in the panel. By dragging more entities to the map, users can search for information related to up to six entities for each cell. Users can freely drag entities to other parts of the map and can thus explore the information space in multiple directions.

InnovationMap does not include methods to share specific entities or entity structures with an other users. However, it is designed to remotely save the state of each session for later access, potentially on another device. As a result, even without features designed specifically for collaboration, sharing access to a session with another user is possible, thus allowing users to provide others with the information they have found, including the context of their multiple search directions and points of interests.

Scenario

An innovation manager for a foreign company is looking for partnership opportunities in Helsinki. She uses InnovationMap, an online tool for exploring research work, provided by the local university. Although she does not have academic expertise, she starts by inputting an umbrella term that roughly covers her fields of interests (i.e., "Big Data"). The system returns a variety of related fields, as well as keywords, academic articles, patents, and names of researchers and entrepreneurs, all of which are representative of the local research work. She consults the profile of some of the recommended persons and inspects a few recommended fields. Soon, she's able to drag "Human Computer Interaction (HCI)" towards the map, thus refreshing the search results. She then adds an article that describes a user-modeling technique for search systems. The results now show a startup based on the technology of interest, among other related elements. Inspection of the entity provides necessary information on the young company, including a link to its website and contact information. In parallel, she drags "Information Visualization" to the map, next to "Human Computer Interaction (HCI)", and starts exploring the intersection of these topics in the information space (in a separate cell). Progressively, and without having to recall any technical terminology or read through pages of material unrelated to the topic of interest, she not only collects useful information but also builds a representation of the whole process, thus gaining an understanding of the information space that accurately reflects the data.

DISCUSSION AND CONCLUSIONS

The present template consists of fundamental principles aimed at guiding the design of future systems and supporting information exploration while also limiting the number of constraints imposed on the overall design space. In this section we discuss aspects that are not addressed by the template, and attempt to outline the remaining design space.

Hypercues are designed to be identified and defined by users (although they could also be recommended within contents).

For instance, in the latest iteration of its operating system for tablets (iOS 11), Apple has introduced a generalized ability to drag and drop. Pictures, text snippets, news articles, hyperlinks, and other bit of information pop out of the environment with a gesture of the finger, thus becoming interactive objects that can be dragged across applications and dropped into messages, notes, or cloud-based storage. This ability lets the user interact with predefined object and with user-defined selections, and it offers an ideal interactive base for the integration of the affordances proposed in this paper.

Although the template does not provide information about the shape and size of displayed hypercues, it is useful to discuss the requirements and provide some recommendations based on our experience. The first requirements of the hypercue marker is for the represented entity to be identifiable and placed in a space that allows it to be moved and positioned in relation to other entities. A constant challenge when designing entity-based interfaces balancing the amount of information conveyed through the entity marker against the number of entities that can be comfortably displayed. In any case, it is necessary to provide the option to quickly inspect the entity, so that the user can access a comprehensive overview of the entity through linked content and related material. However, it is also essential to show enough information up-front to trigger the user's recognition and incite her interest. Modern desktop-based operating systems offer a good model for representing files and directories as manipulable objects using an icon and one or two short lines of text. The most useful information depends on the task and on the information space. Although movies are usually displayed with a poster, a title and a release year, finding the most relevant movie in a set could depend on other information, such as the cast or the rating. Likewise, finding useful academic articles can require variable criteria (e.g., authors, venue or citations). The solution might lie in a balance between user-defined preferences and automated, context-sensitive, and adaptive interface settings.

These guidelines' reliance on direct manipulation and spatial layouts makes the hypercue a potentially interesting candidate for integration with the physical world through playful tangible interactions. Registering an entity or a set of entities as physical objects allows users to combine and share such objects to playfully discover information through machine vision or sensing surfaces.

The hypercue template consists of a minimal set of affordances for the interactive representation of real-world entities. It ensures that all important features for supporting exploratory search can be addressed while still leaving enough design space to facilitate integration within a variety of systems, and it baseline rules and expectations to facilitate users' engagement in complex search behavior. The template has various potential implications regarding how people search for information. An important implication is that all interactions proposed in the present work require substantially more effort from the user than present methods require, as users have grown accustomed to content feeds and to the simplicity and immediacy of today's search engines. We advocate information practices in which users are more active, and we posit that this is the cost of providing greater transparency and control over information. However, this cost can be mitigated through fluidity by having every interaction serve an informational goal and letting the user become truly absorbed by the task, thus rewarding her with persistent and constructive search sessions that remain useful in the long run.

In this work, we propose guidelines that are generalizable to every information space. We can imagine that, in the future, a user's search interface of choice will be independent from the data being explored. As a result, a user could apply one tool to discover information of interest within news posts, academic articles, music, movies, and social media posts, thus increasing the potential for serendipity and creative solutions.

ACKNOWLEDGMENTS

This work has been supported by the Academy of Finland (D4Health 295506, 313610 and 312274). The data used in the experiments is derived from the ACM Digital Library. We would like to thank Luana Micallef for helpful discussions and valuable suggestions.

REFERENCES

1. Christopher Ahlberg and Ben Shneiderman. 1994. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. ACM, New York, NY, USA, 313–317. DOI: http://dx.doi.org/10.1145/191666.191775

2. Saleema Amershi and Meredith Ringel Morris. 2008. CoSearch: a system for co-located collaborative web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1647–1656.

3. Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* 13, 5 (1989), 407–424.

4. Eric A Bier, Stuart K Card, and John W Bodnar. 2008. Entity-based collaboration tools for intelligence analysis. In *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*. IEEE, 99–106.

5. David Carmel, Naama Zwerdling, and Sivan Yogev. 2012. Entity oriented search and exploration for cultural heritage collections: the EU cultura project. In *Proceedings of the 21st International Conference on World Wide Web*. ACM, 227–230.

6. Marian Dörk, Sheelagh Carpendale, and Carey Williamson. 2011. The information flaneur: A fresh look at information seeking. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1215–1224.

7. Niklas Elmqvist, Andrew Vande Moere, Hans-Christian Jetter, Daniel Cernea, Harald Reiterer, and TJ Jankun-Kelly. 2011. Fluid interaction for information visualization. *Information Visualization* 10, 4 (2011), 327–340.

8. Dorota Głowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Samuel Kaski, Giulio Jacucci, and others. 2013. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 117–128.

9. Gene Golovchinsky, John Adcock, Jeremy Pickens, Pernilla Qvarfordt, and Maribeth Back. 2008. Cerchiamo: a collaborative exploratory search tool. *Proceedings of Computer Supported Cooperative Work (CSCW)* (2008), 8–12.

10. François Guimbretière, Maureen Stone, and Terry Winograd. 2001. Fluid Interaction with High-resolution Wall-size Displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology (UIST '01)*. ACM, New York, NY, USA, 21–30. DOI:http://dx.doi.org/10.1145/502348.502353

11. Marti A Hearst. 2014. What's missing from collaborative search? *Computer* 47, 3 (2014), 58–61.

12. Samuli Hemminki, Kai Zhao, Aaron Yi Ding, Martti Rannanjärvi, Sasu Tarkoma, and Petteri Nurmi. 2013. Cosense: A collaborative sensing platform for mobile devices. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 34.

13. Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. 1985. 1.1 direct manipulation: astep beyond programming languages. *Human-Computer Interaction* 1, 4 (1985), 311–338. DOI: http://dx.doi.org/10.1207/s15327051hci0104_2

14. Andruid Kerne and Steven M Smith. 2004. The information discovery framework. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. ACM, 357–360.

15. Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4189–4198.

16. Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. 2017. Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 57–66.

17. Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (2006), 41–46.

18. Iris Miliaraki, Roi Blanco, and Mounia Lalmas. 2015. From "Selena Gomez" to "Marlon Brando": Understanding Explorative Entity Search. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 765–775. DOI: http://dx.doi.org/10.1145/2736277.2741284

19. Meredith Ringel Morris and Eric Horvitz. 2007. SearchTogether: an interface for collaborative web search. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*. ACM, 3–12.

20. Jeanne Nakamura and Mihaly Csikszentmihalyi. 2014. *The Concept of Flow*. Springer Netherlands, Dordrecht, 239–263. DOI: http://dx.doi.org/10.1007/978-94-017-9088-8_16

21. Don Norman. 2013. *The design of everyday things: Revised and expanded edition*. Basic Books (AZ).

22. Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.

23. Peter Pirolli and Stuart Card. 1999. Information foraging. *Psychological review* 106, 4 (1999), 643.

24. Jeffrey Pound, Peter Mika, and Hugo Zaragoza. 2010. Ad-hoc Object Retrieval in the Web of Data. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*. ACM, New York, NY, USA, 771–780. DOI: http://dx.doi.org/10.1145/1772690.1772769

25. Gonzalo Ramos and Ravin Balakrishnan. 2003. Fluid Interaction Techniques for the Control and Annotation of Digital Video. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 105–114. DOI: http://dx.doi.org/10.1145/964696.964708

26. Tuukka Ruotsalo, Kumaripaba Athukorala, Dorota Głowacka, Ksenia Konyushkova, Antti Oulasvirta, Samuli Kaipiainen, Samuel Kaski, and Giulio Jacucci. 2013. Supporting exploratory search tasks with interactive user modeling. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–10.

27. Daniel M Russell, Mark J Stefik, Peter Pirolli, and Stuart K Card. 1993. The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*. ACM, 269–276.

28. Ben Shneiderman. 1993. 1.1 direct manipulation: astep beyond programming languages. *Sparks of innovation in human-computer interaction* (1993), 17.

29. Ben Shneiderman. 1997. Direct manipulation for comprehensible, predictable and controllable user interfaces. In *Proceedings of the 2nd international conference on Intelligent user interfaces*. ACM, 33–39.

30. John Stasko, Carsten Görg, and Zhicheng Liu. 2008. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization* 7, 2 (2008), 118–132.

31. Ryen W White and Resa A Roth. 2009. Exploratory search: beyond the query-response paradigm (Synthesis lectures on information concepts, retrieval & services). *Morgan and Claypool Publishers* 3 (2009).

Aalto-DD 192/2019

ISBN 978-952-60-8774-0 (printed)
ISBN 978-952-60-8775-7 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Arts, Design and Architecture
Department of Media, Aalto Media Lab
www.aalto.fi

BUSINESS +
ECONOMY

ART +
DESIGN +
ARCHITECTURE

SCIENCE +
TECHNOLOGY

CROSSOVER

DOCTORAL
DISSERTATIONS