UNDERSTANDING AND SUPPORTING INFORMATION SEEKING TASKS ACROSS MULTIPLE SOURCES

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus Prof. dr. D. C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 8 december 2010, te 10:00 uur

door

Alia Khairia Amin

geboren te Bandung, Indonesië.

Promotiecommissie

Promotor: Prof. dr. Lynda Hardman
Copromotor: Dr. Jacco van Ossenbruggen

Overige Leden:

Dr. Vanessa Evers
Diane Kelly, PhD.
Prof. Alan F. Smeaton, M.Sc., PhD., F.I.C.S.
Prof. dr. ir. Arnold Smeulders
Prof. dr. Bob J. Wielinga

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

SIKS Dissertation Series No. 2010-51

The research reported in this thesis has been carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

This research was supported by the MultimediaN project through the BSIK program of the Dutch Government and by the European Commission contract FP6-027026, KSpace.

Contents

Acknowledgement

Lynda, my wise professor, said a PhD student's journey is similar to that of Frodo's in Lord of the Rings. Sure enough, there have been many challenges faced and conquered during this journey. To say this thesis is solely a result of my personal expedition overstates the case. This work would not exist without the direct and indirect support of many people.

First on the list are the people with whom I worked. My PhD research could not have taken on the form that is has without the support of my supervisors: Lynda Hardman and Jacco van Ossenbruggen. Lynda and Jacco believed in this research even at times when I had doubts. I am in debt to Jacco for many technical insights and his pragmatic approach to finding solutions. Lynda kept me focussed on my work when I started going astray. They kept me on track with my research. We did not always have the same ideas about the work, but the best thing about our three way collaboration was that one of them was always on my side if we had a difference in opinion. I will always remember the funny but accurate metaphors of Lynda and Jacco during our discussions, such as *"Do research like the Dutch women's volleyball team: just focus on having fun"*, *"Correcting a paper is like combing lice from one's hair"* or *"We want you to go to the moon while you just want to cross the street"*. Another fellow researcher with whom I interacted the most was Michiel Hildebrand. I could not have asked for a better partner for PhD research than Michiel. Many seemingly impossible ideas were made possible through his magical Prolog code. In social events, his endless energy at parties and his cooking skills are hard to match.

I am thankful to colleagues at CWI (Corrado, Henning, Theodora, Raphaël, Željko, Joost, Katya, André, Vera, Stefano, Lloyd, Frank, Steven, Michiel Kauw-A-Tjoe, George, Jop, Matthieu, Stefani and José) for all the juggling, salsa dancing, chocolate eating and other fun activities. I would also like to thank Hyowon Lee, a great designer, for many insightful discussions and his friendship. A special thanks to Junte Zhang whose hard work is reflected in Chapter 6 of this thesis, and Niels Nes and Arjen de Rijke who have been a great help on technical hardware issues especially during my last years.

The work presented in this thesis was funded by the E-Culture MultimediaN project. I would like to thank the past project team members (Borys, Lora, Bob, Jan, Anna, Victor, Mark, Laura, Zhisheng) for their cooperation and collaboration with me. I would like to thank Guus Schreiber, for giving user research a special place in a technology-driven project.

I am very fortunate to have worked with a group of very talented people at Google (Sian Townsend, Liviu Tancau and Terry Van Belle). It was a wonderful and empowering experience. I am also very happy to work with great and supportive colleagues at Elsevier. Thank you for making hard work fun.

I am grateful to Alan Smeaton, Diane Kelly, Arnold Smeulders, Vanessa Evers, and Bob Wielinga for serving as members of my PhD committee and for the feedback on my manuscript.

To my brother and sisters, thank you for supporting me with what I do. To my childhood friends and adopted family, Carolina, Nonnie, Jeany, Siska, Archi and Taniatje, I thank you for many years of laughter, for your friendship and for keeping me grounded. To Csaba, thank you for being patient with me, for understanding what is important to me, and for finding common and enjoyable paths for both of us to walk on.

Finally, the two people most responsible for who I am are my parents. Their biggest gift to me is the freedom to pursue the things I love doing. Terima kasih atas segala doa dan restu yang telah diberikan selama bertahun-tahun. Mama, this degree is yours too.

CHAPTER 1

Introduction

People look for information in a variety of ways. Sometimes we search for answers to specific questions on the web while at other times we keep up-to-date by reading information from newspapers. The notion of information seeking is an important part of our daily routine. In parallel, the number of online information sources is growing rapidly. To address this, both academia and industry are active in the research and development of new information access tools. This is evident by the number of new search engines introduced every year offering to help users search for information located in multiple sources on the Web[1]. However, the user interfaces of these new search tools have changed little since they were first introduced more than a decade and a half ago. Most search tools use a single search box as their primary user interface. This raises the question of whether current interfaces and hence search engines provide sufficient support for the wide range of information seeking tasks.

The work described in this thesis addresses this question. Our work examines different information seeking tasks influenced by different user types, different domains and environments. Our aims are to:
a) better understand users' information seeking tasks,
b) identify specific users' requirements while searching for information,
c) support users by providing interface features that satisfy needs that are not yet supported by current tools.

This chapter is organized as follows. In the next section we discuss the two main concepts of this research as background information (section 1.1 and 1.2). Afterwards, we introduce the three research questions of our research (section 1.3). This is followed by explanations of the scope of this thesis (section 1.4). Afterwards, we provide short summaries of the chapters of our thesis (section 1.5) and establish

[1] http://en.wikipedia.org/wiki/Web_search_engine

the important contributions of our research (section 1.6). Finally, we list the publications on which the thesis is based (section 1.7).

1.1 Information seeking tasks

We adopted the description from (Wilson 2000) and defined information seeking task as: *the act of seeking information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems, such as a newspaper or a library, or with computer-based systems, such as the World Wide Web.* Search behavior has been studied extensively, e.g. in (Case 2006; Fisher et al. 2005). In principle, there are many ways to classify and explain users' behavior while searching for information. Most of these models fall into one of two types:
a) *Holistic models:* Generalizations or principles of users' information seeking behavior explained as complex interactions between internal (cognitive, affective) and external (environment, work) factors. For example, Wilson's information behavior model describes the interaction between the environment, the person and social role (Wilson 1999), or the Ingwersen and Järvelin's general model of cognitive information seeking and retrieval describing the interaction between information objects (text, images), the IR system, the interface, the cognitive actors and social/organizational/cultural context (Ingwersen and Järvelin 2005).
b) *Empirical models:* Models of users' information seeking behavior are derived from observations. The research tends to explain information seeking from a specific aspect, such as by users' goal (Kellar et al. 2007; Sellen et al. 2002), by search process (Choo et al. 2000a; Kuhlthau 1991), by the way users use certain information access tools (Kelly and Belkin 2001; Rose and Levinson 2004; Jansen et al. 2007), or by the type of content people search for (Yi et al. 2008).

In our research, we adopt the empirical goal-based information seeking task taxonomy as the base model to explain users' information seeking behavior for two reasons: first, the validity of the taxonomy can be empirically re-examined in different settings (e.g. different user profiles, different domains and different tools). Second, the taxonomy should explain users' intentions independent of what the current tools allow them to do. The goal-based information seeking taxonomy has been investigated and verified by previous research. Table 1.1 shows related research on goal-based information seeking tasks. Even though using different terminologies, they reflect similar ideas on information seeking task categories:

- Fact finding task: users have a specific goal and focused questions — they typically look for specific factual pieces of information.

- Information gathering task: users carry out several search tasks to fulfill a higher level goal, such as collecting information to write a report.
- Non-goal oriented task: users' tasks are not goal driven, but rather "keeping up-to-date", "just browsing", "serendipitous discovery" or "see what's new or interesting".

Table 1.1: Comparison of goal-based categories of information seeking tasks

	Information Seeking Tasks Category		
Research	fact finding	information gathering	non-goal oriented
Choo (2000) & Marchonini (1995)	informal search	formal search	undirected viewing
Morrison et al. (2001)	find	compare, collect understand, explore	monitor
Sellen et al. (2002)	finding	information gathering	browsing
Marchonini et. al. (2006)	look up	learn, investigate	-
Kellar et al. (2007)	fact finding	information gathering	just browsing

User type, situation and environment can shape the way users perform information seeking tasks. Thus, when designing an application to support a specific task, it is important to understand the user and the context behind the task. On the flip-side of the coin, we also need to understand what information is available and where it comes from. The information that users need may come from multiple sources. Thus, how to present information from multiple sources is another important aspect in designing an information access application.

1.2 Searching across Multiple Sources

Historically, research on integrated services for multiple information sources is technology-centered. One goal of this research is typically to link and merge the different sources so that users can access them as a single platform. It has been acknowledged as an important and challenging research topic for many communities, such as digital libraries, e.g. (Endig et al. 2000; Pyrounakis et al. 2004), information retrieval, e.g. (Si and Callan 2005), databases, e.g. (Manolescu et al. 2001) and semantic web, e.g. (Schreiber et al. 2008). These communities primarily discuss technical issues of multiple source search, e.g. merging techniques, interoperability, query efficiency or ranking.

Previous research has not considered searching across multiple sources from the end-user perspective. With the ability to search across multiple sources, however, new challenges and problems arise. For example, how should (combined) information from multiple sources be presented in a way that a user can understand the difference or how can users navigate through multiple (unfamiliar) information sources. End-user search across multiple sources is acknowledged to be a nontrivial problem by the human computer interaction community, e.g. (Aula and Russell 2009; Baldonado 2000). Searching across multiple sources remains an HCI topic that has yet to be fully explored and understood.

In this thesis, we follow a user-centered design approach to find novel solutions to support information seeking tasks across multiple sources. This translates to three steps of research: first, understanding users' information seeking tasks; second, deriving requirements to support different information seeking tasks; last but not least, exploring different means to support information seeking tasks and evaluate whether the solution fulfills the initial requirements.

1.3 Research questions

The three steps mentioned above are the basis of our research questions. The first research question is dedicated towards understanding different users' information seeking tasks. Different users may have different information seeking tasks. What these tasks are and which tasks lack support tools need to be clearly identified. The first research question, therefore, is:

Research Question 1. What are the information seeking tasks for cultural heritage experts and mobile lay users?

Specifically, we investigate the different types of information seeking tasks across multiple sources for experts and lay users, taking examples from the cultural heritage and location-based mobile search application domains.

The answers from the first research question provide us with a better understanding of users' information seeking tasks. Moreover, they give us insights on features of information access tools that are suitable for supporting users' tasks. These insights are used as input to derive user requirements for information access tools. This is the focus of our second research question:

Research Question 2. What are the requirements to support information seeking across multiple sources?

These requirements build on our findings from the first research question. They cover functional and user interface requirements to design future applications for experts and lay users.

Finally, based on the list of requirements provided by the second research question, we explore different interface and interaction features to support specific information seeking tasks. This is the focus of the third research question:

Research Question 3. How can we support information seeking tasks across multiple sources?

We take several requirements from the second research question and use them as guidance in designing information access tools. To further validate our requirements, we conduct user evaluations for these tools.

In order to be able to answer the research questions, we need to investigate specific instances of information seeking tasks performed by specific users in specific application domain. We also need to select and focus on technical solutions. This is discussed in the following section.

1.4 Scope

In order to have a more comprehensive view on the different information seeking tasks, we have chosen to investigate two user profiles: experts users (chapters 2 and 5) and lay users (chapters 3, 4 and 6). Lay users have no prior knowledge of the domain being used. Experts have some level of expertise of a particular domain, conduct search regularly for their professional work and have some understanding of the quality of information sources used.

We have chosen to investigate two different domains: cultural heritage domain (chapters 2, 5 and 6) and location-based and geographic domain (chapters 3 and 4). We concentrate on the cultural heritage domain because, within the context of the MultimediaN E-Culture project[2], we had access to a subset of cultural heritage and geographic datasets and thesauri. The cultural heritage domain is a knowledge-intensive domain, characterized by rich and heterogeneous data types from large information sources, while the location-based domain tends to use common knowledge, e.g. the Web.

Due to the differences in the domain familiarity and the context of the information seeking tasks between these two user groups in two different domains, we

[2] http://e-culture.multimedian.nl/

expect to get a wider overview of the different instances of information seeking tasks.

We acknowledge that while different technical solutions may be required to solve different kinds of information seeking task challenges, in this research, we focus on thesauri to link information stored in multiple cultural heritage information sources. This approach is taken for a pragmatic reason: museums, libraries and other cultural heritage organizations have a long history of carefully annotating the objects in their collections. Many cultural heritage institutes make use of thesauri for their annotations. They use thesauri to limit the problems related to multiple terms, such as the use of synonyms, name variants, differences in languages or differences in jargon, that occur within their own collections. This research was conducted within the context of the MultimediaN E-Culture project. Within the project, we have investigated to what extent these thesauri can also be used to address the same problems when searching across multiple sources (Schreiber et al. 2008).

1.5 Thesis overview

This thesis is divided into two parts. The first part of the thesis, chapters 2 and 3, consists of research that provide us with understanding of the issues concerning information seeking tasks for experts and lay users (problem oriented research). The second part of the thesis, chapters 4, 5 and 6, consists of research that focuses on specific information seeking task problems and explores alternative interfaces to resolve these problems (solution oriented research).

Chapter 2 presents a user study on investigating information seeking behavior of cultural heritage experts (Amin et al. 2008). We explain in detail the different information seeking tasks that occur within the cultural heritage domain (research question 1). Experts from different cultural heritage institutions participated in semi-structured interviews to explain their information seeking tasks as part of their daily work. The study provides insights on why and where experts search. Furthermore, the study reveals positive and negative aspects of the state of the art search tools used by the experts (research question 2).

Chapter 3 describes a digital diary study on investigating location-based information seeking behavior for lay users in a mobile context (Amin et al. 2009). The study focuses on what kinds of day-to-day mobile information seeking tasks occur while people are on the move (research question 1). The results indicate typical location-based mobile information seeking tasks, the temporal, spatial and social context in which these tasks occur and how the context influences people's search tasks and decision making. Finally, we discuss recommendations on how to improve services for such tasks (research question 2).

The use cases and results of the behavioral studies from the first part of the thesis are used as inspiration for requirements of the future information access tools. The following chapters discusses the design rationale and evaluation of these tools (chapter 4, 5 and 6). Since the functionalities offered by these tools are made possible with thesauri, we call these applications thesaurus-based applications.

Chapter 4 presents two user interface evaluations on the use of thesauri to help lay users find the correct keywords in an autocompletion interface (Amin et al. 2009; Amin et al. 2008; Hildebrand et al. 2007). The aim of the study is to derive guidelines and recommendations for developing thesaurus-based autocompletion interfaces for query formulation (research question 2). Several variations of thesaurus-based autocompletion interfaces using two thesauri were evaluated. The evaluation provides us with insights into how to use thesaurus-based autocompletion to support query formulation (research question 3).

Chapter 5 discusses the user requirements, design and evaluation of a thesaurus-based comparison search application (Amin et al. 2010). We discuss requirements for interfaces to support comparison search in the cultural heritage domain (research question 2). The application enables experts to compare multiple artworks simultaneously. In the evaluation, we compared two applications, the thesaurus-based application and the application the users currently use, and investigate how well the two applications support different aspects of comparison search. The results of the evaluation provide recommendations on how to support comparison search tasks for cultural heritage experts (research question 3).

Chapter 6 describes an exploration study on the effects of information source credibility ratings in an information aggregator (Amin et al. 2009; Zhang et al. 2009). We focus on investigating how visualization influences lay users' confidence while accessing information from different cultural heritage information sources (research question 3). When a user conducts an information seeking task, s/he also needs to decide which information sources to rely on. Consequently, it is important for the user to be confident about the credibility of the information source. Our user evaluation shows how users' confidence can be influenced when information source credibility ratings are shown.

Chapter 7 describes answers to the research questions. We identify the information seeking tasks for cultural heritage experts and mobile lay users. We discuss the functional and user interface requirements for future information access tools and discuss several interface features to support specific information seeking tasks. Moreover, we reflect on the challenges faced when evaluating applications for multiple sources and discuss the limitation of our research. Finally, we provide directions for future work in this area.

1.6 Thesis contributions

The work presented in this thesis contributes to the research and development of information seeking across multiple sources. We provide:

- An in-depth analysis of information seeking tasks in multiple information sources for experts and lay users (research question 1). The analysis enriches previous theory on information seeking tasks. For example, we improve the information seeking task taxonomy by identifing the different types of information gathering tasks. We identify the general trends in experts' information seeking behavior. Moreover, we further clarify the information seeking behavior for experts and lay users with respect to the information sources used, the social, temporal and spatial context behind information seeking tasks, and the challenges people face when conducting these tasks.

- A list of functional and user interface requirements for future information access tools (research question 2). We translated these requirements to a set of design recommendations and guidelines for interface and interaction design for future information access tools to support information seeking across multiple sources. Furthermore, some of these design recommendations were verified by user evaluations of new interface features.

- Evaluations of the prototype implementation to support the different information seeking tasks (research question 3). Based on our user evaluations, we develop interface design guidelines for the information access tools and derive requirements for the data and technology to implement such tools.

- Identify challenges in evaluating information access tools for multiple sources. We reflect on the user interface evaluation methods used throughout the research and identify interface evaluation challenges related to the dataset and the search algorithm. Moreover, we propose future work to improve the evaluation method for similar tools.

1.7 Thesis material

The material used in the chapters in this PhD thesis is based on the research described in the following publications.

- **Chapter 2.** A. Amin, J. van Ossenbruggen, L. Hardman, and A. van Nispen. Understanding cultural heritage experts' information seeking needs. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, pp. 39–47, New York, NY, USA, 2008, ACM Press.

- **Chapter 3.** A. Amin, S. Townsend, J. van Ossenbruggen, and L. Hardman. Fancy a drink in canary wharf?": a user study on location-based mobile search. In *Interact '09: 12th IFIP TC13 Conference in Human-Computer Interaction*, Uppsala, Sweden, pp.736–749, 2009, ACM Press.

- **Chapter 4.** A. Amin, M. Hildebrand, J. van Ossenbruggen, V. Evers, and L. Hardman. Organizing Suggestions In Autocompletion Interfaces. In *ECIR'09: Proceedings of the Annual European Conference on Information Retrieval*, pp. 521–529, 2009, ACM Press.

- **Chapter 5.** A. Amin, M. Hildebrand, J. van Ossenbruggen, and L. Hardman. Designing a thesaurus-based comparison search interface for linked cultural heritage sources. In *IUI'10: Proceedings of the International Conference on Intelligent User Interfaces*, pp.249–258, 2009, ACM Press.

- **Chapter 6.** A. Amin, J. Zhang, H. Cramer, L. Hardman, and V. Evers. The effects of source credibility ratings in a cultural heritage information aggregator. In *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web*, WWW'09, 35–42, New York, NY, USA, 2009, ACM Press.

- **Chapter 7.** J. van Ossenbruggen, A. Amin, and M. Hildebrand. Why evaluating semantic web applications is difficult. In *SWUI'08: Proceedings of the workshop on semantic web user interface*, CHI'08, Florence, Italy, 2008.

Additional publications by the author are:

- A. Amin. Establishing requirements for information gathering. ECDL'08 Doctoral Consortium. *Bulletin of IEEE Technical Committee on Digital Libraries (TCDL)*, 5(2), 2009.

- L. Hardman, J. van Ossenbruggen, R. Troncy, A. Amin, and M. Hildebrand. Interactive information access on the web of data. In *WebSci'09: Society On-Line*, 2009.

- M. Hildebrand, J. R. van Ossenbruggen, A. K. Amin, L. Aroyo, J. Wielemaker, and L. Hardman. The Design Space Of A Configurable Autocompletion Component. *Technical Report INS-E0708*, Centrum Wiskunde & Informatica, 2007.

- G. Schreiber, A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Ossenbruggen, A. Tordai, J. Wielemaker, and B. J. Wielinga. Semantic annotation and search of cultural-heritage collections: The multimedian e-culture demonstrator. *J. Web Sem.*, 6(4), pp.243–249, 2008.

- G. Schreiber, A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga. Multimedian e-culture demonstrator. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, The Semantic Web - ISWC 2006, vol. 4273 of LNCS, pp. 951–958. Springer, 2006.

- J. van Ossenbruggen, A. Amin, L. Hardman, M. Hildebrand, M. van Assem, B. Omelayenko, G. Schreiber, A. Tordai, V. de Boer, B. Wielinga, J. Wielemaker, M. de Niet, J. Takema, M.-F. van Orsouw, and A. Teesing. Searching and annotating virtual heritage collections with semantic-web techniques. In *Museum on the Web'07: International conference for culture and heritage on-line*, 2007.

- P. Wilkins, R. Troncy, M. Halvey, D. Byrne, A. Amin, P. Punitha, A. F. Smeaton, and R. Villa. User variance and its impact on video retrieval benchmarking. In *CIVR '09: Proceeding of the ACM International Conference on Image and Video Retrieval*, New York, NY, USA, 2009.

- J. Zhang, A. Amin, H. Cramer, L. Hardman, and V. Evers. Improvings user's confidence in cultural heritage aggregated results. Poster paper. In *SIGIR '09: Proceedings of the 32th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, 2009.

Part I

Understanding Users' Information Seeking Tasks

CHAPTER 2

Experts' Information Needs: A User Study on Cultural Heritage Experts

We investigate the information seeking tasks of experts. We focus on experts' tasks using online and offline sources in the cultural heritage domain. The aim of the study is to better understand experts' search motivation, the types of search tools that they use and the challenges that they face while searching (research question 1). The study shows that of the three information seeking task categories, fact finding tasks, information gathering tasks, and non-goal oriented tasks, the majority of experts' search tasks involve relatively complex information gathering. The study also reveals challenges in information seeking, for example while information gathering is an important task for experts, most of their search tools tend to support only fact finding (research question 2). Our observations indicate that there is a mismatch between the experts' information seeking tasks and the available support. Additionally, we discovered that experts search in many different online and offline sources and that one of the challenges in searching through multiple sources is the issue of identifying the credibility of the information sources.

This chapter was published as "Understanding Cultural Heritage Experts' Information Seeking Tasks" in the Proceedings of the Joint Conference of Digital Library (JCDL'08) (Amin et al. 2008) and was co-authored by Jacco van Ossenbruggen, Lynda Hardman and Annelies van Nispen.

2.1 Introduction

Cultural heritage is a vast domain consisting of museums, archives, libraries and (non)government institutions. Searching for information in this domain is often challenging because the sources are rich and heterogeneous, combining highly structured, semi-structured and unstructured information, combining authorized and unauthorized sources, and combining both text and other media (e.g. image and video). To perform their daily work, domain experts need to access and exploit cultural heritage information in its full richness. The specific information seeking needs of these experts remains, however, a scantly researched area. This chapter reports on a user study that was motivated by the need to understand the information seeking behavior of cultural heritage experts. The results of this study are currently being used to improve the search tools developed in the context of the MultimediaN e-culture project (Schreiber et al. 2006).

Key findings of our study include, first, that experts' daily search tasks are dominated by a range of different (relatively complex and high level) information gathering tasks, while the tools tend to be geared towards support for (relatively simple and low level) fact finding tasks. Second, many search tasks require experts to use and combine results from multiple sources, while the tools typically provide access to a single source. Third, we found that direct communication as a means for information transfer is greatly valued by experts. Finally, we found that trust in the information source is an important aspect of experts' search activity.

The chapter is structured as follows. After discussing related work, we describe the setup and analyze the results of the user study. The analysis includes an information task classification of the use cases reported by the participants, a classification of the type of information sources used in these use cases and an analysis of the underlying search tasks. We then discuss the extent to which current tools support the experts' search tasks and illustrate inadequate tool support with concrete examples given by experts during the interviews. Finally, we discuss design considerations based upon our experiences within the MultimediaN project.

2.2 Related work

The motivations behind searching information have been studied extensively (Broder 2002; Case 2006; Choo et al. 2000b; Pirolli 2007; Rose and Levinson 2004). Research by Broder (Broder 2002), extended by Rose et al. (Rose and Levinson 2004), found search motivations such as *navigational search, informational search* or *resource finding*. Their research is mainly based on analyzing logs when people use a search engine and a short survey. It is, thus, difficult to know the real search motivation. Choo et al. (Choo et al. 2000b) took a different approach

and monitored web browsing activities of people for two weeks and conducted an interview to check the participants' search motivations. They found that people have different modes of searching, where each mode has its own traits and search strategy (Choo et al. 1998).
Kellar et al. (Kellar et al. 2005) compared previous research on information seeking task categories (Choo et al. 2000b; Morrison et al. 2001; Sellen et al. 2002) and proposed a taxonomy which gives a more thorough overview on the information tasks. They used this model to explain peoples' behavior on the web. In this study, we extended this taxonomy to discuss online as well as offline information task behavior for expert users. We use the following information tasks categories (adapted from (Kellar et al. 2007; Kellar et al. 2005; Kellar et al. 2006), see Figure 2.1):

1. Fact Finding: users ask goal-oriented and focused questions; they look for specific factual pieces of information.

2. Information Gathering: users carry out several search tasks to fulfill a higher level goal, such as writing a report, preparing an event, or collecting information to make a decision.

3. Non-goal-oriented: this search task is generally not goal-driven, other than to "keep up-to-date", "just browsing", "see what is new or interesting", recreational searching or even "passing time".

4. Communication: an information exchange task, either face-to-face or through technology, such as email.

5. Transaction: an information exchange task, e.g., online auction, banking, downloading multimedia documents. Transactions are often associated with a user name and password combination.

6. Maintenance: a task which involves organizing information, e.g., updating bookmarks or organizing email in the appropriate folder.

To date, research on cultural heritage experts' search behavior is limited. A survey with 477 cultural heritage experts in the Netherlands on Collectiewijzer[1] (van der Graaf 2006) usage reveals that experts think the Internet is becoming a more important information source. Experts say that they would use such system mainly to: do research, compare collections and look for potential items to borrow

[1] Collectiewijzer is an online portal that supports information linking and exchange between cultural heritage institutions

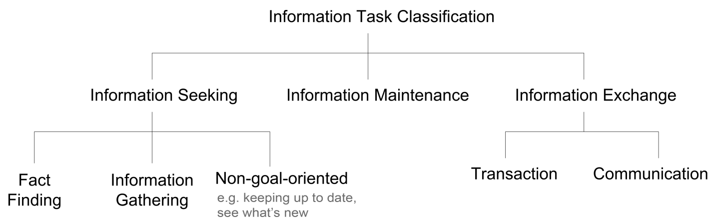


Figure 2.1: Classification of Information Tasks, adapted from (Kellar et al. 2007).

from other collections. The survey suggests that information seeking is an important part of experts' work. Unfortunately, it does not provide insights into experts' search behavior nor the kind of difficulties they experience when searching.

Table 2.1: Participants' Demography (total:17 people)

Digital library experience:*)	expert(3), intermediate(12), basic(2)
Cultural heritage sub-domain:	ethnography(9), Dutch classic art(6), contemporary art(2)
Age: years old	21-30 (3), 31-40 (6), 41-50 (3), 51-60 (5)
Expert role: **)	researcher(6), curator(8), registrar(3), IT(3), teacher(2), student(3)
Affiliation: **)	museum(8), company/freelance(2), university(5), cultural heritage institution(8)

Note:
*) expert: IT experts, also manage the museum information systems.
intermediate: Intensive use of search engines, online and offline digital libraries.
basic: Minimal computer usage, e.g only email and Word. Prefer using traditional libraries.
**) A single expert can have more then one role/affiliation.

2.3 User study setup

The purpose of our study is to understand the search behavior, that is the information seeking tasks, of cultural heritage experts. We, thus, do not investigate other information tasks, such as information exchange and information maintenance. Our main research questions are:

1. *Why do cultural heritage experts search?* What is the motivation that gives context to their search activity?

2. *Where do these experts search?* Which sources do they use, why do they use them and do they experience any problems with them?

3. *What are the experts' search tasks?* What are typical search tasks, which search tasks do they do the most/least and do their tools sufficiently support the tasks?

To answer these questions we conducted a user study, which we describe in the following section.

2.3.1 Procedure

Most of the interviews took place at the participants' working environment. We conducted two pilot studies prior to the actual interviews to make sure all questions would be clearly understood. Each participant was interviewed individually with semi-structured questions and was asked to answer a questionnaire. The interview had three parts:

1. Introduction, demographic questions and informed consent.

2. Questions about the participant's main responsibilities and daily activities at work.

3. For each activity mentioned in part 2, participants were asked to give examples and to describe the purpose of the activity, its frequency and the tools involved. Supplementary questions include their subjective impression based on their experience in using the tools.

After the interview, we asked them to demonstrate the tools they used and to give some examples on how they use them. On average, the whole interview took two to three hours. The interview was voice recorded; pictures and screen shots of the tools were taken. We analyzed the voice recording, photos and screen shots of the experts' tools and questionnaires. Activity descriptions (use cases) for every participant were noted down. Samples of screen shots of the tools from the participants helped us clarify the way our participants carry out their daily work and the problems that they face.

2.3.2 Limitations

We acknowledge that the method used has shortcomings. We did not capture the dynamics of experts' behavior over a longer time frame and relied on participants' (selective) memory. We may thus not have captured unconscious behavior of these experts. This study will also not reveal non work related search such as recreational search. However, we faced several restrictions: many experts are reluctant to give consent on automatic computer monitoring. This is not just because of privacy reasons, but also because it is against institute policy to share sensitive information, such as correspondence between experts, or install unauthorized monitoring software in the organization's computer network. These restrictions are the main reasons behind our pragmatic approach. Despite the approach that we took, we believe the study captured key aspects of the cultural heritage experts' information seeking behavior.

2.3.3 Participants

Based on recommendations from ICN[2], we recruited experts who frequently search for information related to cultural heritage (see Table 2.1). In total, 17 Dutch professionals participated from 9 different cultural heritage institutions in the Netherlands (five museums, two companies, one university and ICN). Most participants use computers intensively at work. We interviewed experts with a variety of backgrounds to capture the different perspectives of information searching needs. Depending on the size of the organization, a single expert can have one or more roles. In large museums, an expert typically has a clearer and more specific role compared with smaller museums where an expert takes responsibility for several roles. We distinguish five expert roles in this study:

Researchers develop guidelines, recommendations, articles and books. Examples of cultural heritage research are developing different conservation techniques or developing theories on the history of acquisition of objects.

Curators are responsible for the collections and their documentation, including arranging loans, acquiring objects and planning exhibitions.

Registrars handle the digitization process of collections in the collection management system. Depending on the size of the museum, the registrars may work together with the curators in annotating collections. Together with curators, they also handle new entries and check whether the information is correct. They also prepare periodical reports on the museum collection status.

Teachers and students were recruited from a relevant Master's program at a Faculty of Art. The students are in their final year and carrying out their internship in

[2] The Netherlands Institute for Cultural Heritage `http://www.icn.nl/`

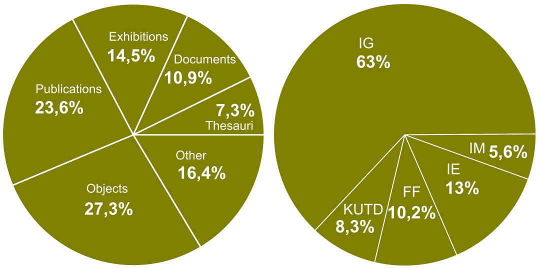


Figure 2.2: Types of experts' use cases (left). Types of experts' information seeking tasks (right).

a museum. The lecturers' main search activity is to prepare their course materials and keep up to date with the state of the art in the field. Students need to search regularly for making reports and assignments for their class. They also assist the researchers, curators and registrars while doing their internship.
IT personnel manage the museum database system. They prepare customized user manuals for the system and they often train and support other employees. In addition to this, they update the museum website, create regular reports on the status of collection documentation and assist registrars on collection documentation.

2.3.4 Classification of use cases

For every participant, we noted their main activities (i.e. use cases), which were further broken down into one or more sub-tasks. We filtered out the use cases that do not involve information seeking activities in the cultural heritage domain, such as project management or fund raising. In total, there are 17 participants with 53 use cases and 110 tasks. We then identified all information seeking tasks occurring within each use case. All information tasks were classified independently into one of the five information task categories by two reviewers. They were guided by information task descriptions similar to those given in the Related Work section. Cohen's kappa was used to measure the agreement between the two reviewers. We found the agreement of $\kappa = 0.74$ (sufficient). The main disagreement between the two reviewers was in deciding which task category was more dominant for a given sub-task. This occurs for complex tasks, such as Information Gathering.

2.4 User study results

This section is divided into three parts, reflecting the three research questions in the user study. First, we discuss the experts' motivation that gives context to why they need to search for cultural heritage information. Second, we discuss the sources that the experts use to find information. Finally, we discuss the different types of experts' search tasks with examples.

2.4.1 Why do experts search?

We classified all use cases into five groups (see also the left of Figure 2.2):
1. Object handling. Experts need to gather information for the restoration, acquisition, loan or sale of an individual object. For example, before acquisition of a painting, a curator needs to research the history of the painting to see whether it is suitable for the museum's collection.
2. Exhibition planning. Experts spend a great deal of effort in research when preparing for an exhibition, e.g. finding interesting themes, carrying out comparison studies with previous exhibitions and publications. Serendipity is highly valued here. The main goal is to find different and interesting perspectives to present to the public.
3. Research for publication. Experts' publication activities can be divided into preparing publications for peer experts or for the general public. The first is mainly about developing guidelines, best practices, recommendations related to the cultural heritage domain and presenting new findings and discoveries. This activity also includes dissemination of the research through lectures and presentations. Publications for the general public are typically PR-related activities, such as writing an *"object of the month"* section for the website for which the history of the object, including interesting facts that could attract the general public, need to be found.
4. Managing the collections' documentation. Records in the collection management system are constantly updated by curators and registrars. For example, new objects need to be registered, information about old collections needs to be updated. When a new object is registered, the expert needs to compare annotations of similar objects and search for more descriptions by looking further in various sources.
5. Building thesauri. Thesauri are controlled vocabularies used, in this case, for annotating objects in museum records. Within an organization, it is important that everyone use the same terms to express the same concepts in the museum records. So experts need to collect terms important for the field from selected sources such as literature, dictionaries, library archives and object descriptions for

Table 2.2: Frequency mention of source (total:204 mentions)

	Source	
1	literature (lit.)	21.4%
2	archives and catalogs	18.9%
3	personal contact	10.2%
4	visit exhibition/museum	1.5%
	Total offline sources:	**5**2.0%
5	reputable website	18.9%
6	ollection management system (CMS)	13.8%
7	search engine	8.7%
8	other digital sources	6.6%
	Total digital sources:	**4**8.0%

Table 2.3: Distribution of sources for different types of use cases (total:204 mentions)

Usecase type vs. source	lit.	reputable website	archives catalogs	search engine	CMS	personal contact	visit exh. museums	other
1.Objects	9	9	11	3	8	4	1	1
2.Exhibitions	9	7	10	4	3	3	1	1
3.Publications	17	10	8	4	7	6	1	8
4.Doc.	4	2	2	1	7	0	0	0
5.Thesauri	4	7	1	2	2	1	0	1
6.Other	3	3	7	3	1	6	0	2
Total	46	38	39	17	28	20	3	13

inclusion in their thesauri. Information from reputable websites, such as those of other museums or cultural heritage organizations, are also used. Typically, multiple experts need to agree on the proposed terms before they are included in the thesaurus. Different cultural heritage branches may use different thesauri. Examples mentioned by our experts are the SVCN[3] thesaurus for Dutch ethnography and the AAT/AATNED[4] for general art and architecture terms.

[3] http://www.svcn.nl/

[4] http://www.getty.edu/research/conducting_research/vocabularies/aat/

Table 2.4: Distribution of the experts' information seeking tasks (total:110 use cases)

Search task vs. Use case type	Total	Thesauri	Doc.	Exh.	Publ.	Obj.	Other
Fact finding (FF)	11 (10.2%)	0	1	0	5	3	2
Inf. Gathering (IG)	70 (63.0%)						
1.Comparison	9	0	2	0	2	3	2
2.Relationship search	6	0	0	2	1	1	2
3.Topic search	39	4	4	7	10	8	6
4.Combination	9	0	1	3	3	2	0
5.Exploration	7	0	0	5	1	1	0
Keeping up to date (KUTD)	9 (8.3%)						
1.Active	7	1	1	0	1	4	0
2.Passive	2	0	0	0	1	0	1
Inf. Maintenance (IM)	6 (5.6%)	0	2	1	2	1	0
Inf. Exchange (IE)	14 (13.0%)						
1.Transaction	2	0	1	1	0	0	0
2.Communication	12	2	0	1	5	2	2

2.4.2 Where do experts search?

Experts consult a large variety of sources to look for the answers they need (see Table 2.2 and 2.3 for details). Source credibility is an important aspect for experts. For research, they carefully select and refer only to reliable sources.

1. Literature: This includes magazines, dictionaries, books, publications, biographies, encyclopedias, journals and reference databases such as RKD library[5] or Picarta[6]. Offline literature remains the most important source for experts because comprehensive knowledge on art, culture and history are usually in books and not yet available in digital form. Online integrated bibliographical search, such as Picarta, helps experts in their search for the correct literature across many libraries.

2. Archives and catalogs: This includes exhibition catalogs, auction logs, inventory cards and remarks fields made by curators. Old documention about objects, often not yet digitalized in the collection management system, are stored as inventory cards and remarks on log books.

3. Personal contact: Personal contact and networking remain one of the most important means of seeking information. For example, communication between

[5] `http://www.rkd.nl/rkddb/`

[6] `http://www.picarta.nl/`

curators is about each other's collection or to find out more about a particular historical or cultural topic.
4. Visit exhibitions or museums: Experts gain knowledge by conducting working visits to other museums to see their collections in person.
5. Reputable Internet sources: It is extremely important for experts to use reliable information sources, including sources on the Internet. The museum curators, who are responsible for the annotation of their objects, agree on which online resources meet their quality standards. Reputable Internet sources are, for example, specific museums websites, cultural heritage institution websites (SVCN website[7], RKD online, galleries.nl), global gazetteer[8] and the CIA fact book[9].
6. Collection management system Each museum maintains its own information system. This stores the records of all objects and is often an integrated system to help museum employees with almost all aspects of their work, such as management, status report generation and loan request processing. Examples of commonly used commercial systems are TMS[10] and ADLIB[11].
7. Web search engines: Most experts mentioned Google as their search engine. In contrast with how experts use reputable websites, experts use search engines for navigational search (e.g., to find an artist, museum or gallery website) and, interestingly, to seek inspiration (e.g. *"what does Google have on Iranian Calligraphy?"*).
8. Other digital sources: This includes all tools not mentioned above, such as online newspapers or RSS feeds.

Table 2.2 shows a summary of the sources used by the participants. Figure 2.3 shows the interfaces to these sources. We can see that the number of use cases using traditional sources (1–4 add up to 52%) and digital sources (5–8 add up to 48%) is comparable. While more and more sources are made available for online search, offline literature remains a very important source of cultural heritage information.

2.4.3 What are the experts' search tasks?

We originally hypothesized that information seeking tasks are dependent on the expert's role. The analysis, however, suggested otherwise: tasks depend more on the type of use case. This is because most of the use cases mentioned by the experts are done in team work by multiple expert roles. Consequently, some of the use cases are overlapping. For example, management of collection documentation is done by

[7] `http://www.svcn.nl/`
[8] `http://www.allm-geodata.com/products1.htm`
[9] `https://www.cia.gov/library/publications/the-world-factbook/`
[10] `http://www.gallerysystems.com/products/tms.html`
[11] `http://www.adlibsoft.com/`

curators and registrars, and student interns help curators with their daily work. We have thus classified the use cases into different task categories of Figure 2.1. The breakdown of each task is summarized in Table 2.4 and the right hand side of Figure 2.2. We discuss each information seeking task: fact finding, information gathering and keeping up-to-date, and illustrate each task with examples given by our participants.

2.4.3.1 Fact finding

Fact Finding search questions vary from simple to very complex. Typical examples of simple queries include:
"What is the contact information for the gallery?" [P4]
"To which tribe/culture does this object belong?" [P1]
"From where does this object originate?" [P3]
Complex queries typically combine several constraints. For example, a curator was given an assignment to select and lend several paintings from the collection for a government building. She needed to select several paintings with appropriate themes and sizes for the building, that would fit on a wall having certain space constraints: *"Are there paintings from our collection, either depicting Amsterdam or created by a painter from Amsterdam, with a width smaller than 50 cm?" [P6]*

2.4.3.2 Information Gathering

With 63%, Information Gathering tasks dominate our expert's use cases. Based on the similarities between the use cases within this group, we identified the following sub-tasks: (See also Table 2.4):
1. Comparison involves gathering information to compare differences and similarities between objects or sets of objects. For example: a curator needs to make an acquisition proposal each year. To do this, she needs to make an assessment of the objects currently in their collection and in that of others: *"What objects from the Middle-East do other museums in the Netherlands have? Is there any tribe or region not represented in our collection or in the collection of other museums? If there is, we need to find out exactly what kind of object we should get." [P14]*
2. Relationship search is about finding relationships between individual pieces of information. For example, a curator needed to research the network of people around the Dutch painter Rembrandt van Rijn. To do this, she performs a literature study, searches for close and distant family members of the artist, the mostly rich and influential people whom he had portrayed, the people whom he had met socially and been friends with. The questions asked about these people are the same every time: *"Who is this person, what does s/he do and what is the nature*

of his/her relationship with Rembrandt?" [P8]
3. Topic search queries can typically be formulated as *"Tell me more about ..."* questions. For example: one of our curator's responsibilities is to maintain the descriptive labels that are associated with all objects in the collection. Among the objects is a specific Jewish ceremonial coat. To determine the history of this coat, the curator checks literature, newspaper archives, auction records, etc.: *"Where and when was this coat made? Was there any restoration done to the coat? What is the purpose of the coat? What does it symbolize? Is there any meaning behind the embroidery? Where was it used? Who used it? Was it ever used in an important historical event?" [P8]*
4. Exploration or exploratory search, is typically not goal-directed. Instead, the expert may associatively follow one train of thought after another. For example, one of our experts was looking for art suitable for decorating the staircase area of a public space. Given that "staircase art" is not an established genre, the expert knows that searching on this term directly will not provide the intended results. Instead, the expert looks for related projects for suggestions, such as artists who do *landscaping* or *city planning art* projects: *"... On specific situations, (such as) in the Staircase project, I look a lot at similar examples of artworks in staircases, for instance, art projects connected to landscaping or city planning, something like that. " [P4]*
5. Combination is about finding matches among pieces of information, most likely from different sources. This task is similar to fitting pieces of a puzzle together to see the bigger picture. For example, a new part of a public building needs to be decorated and the client (the government) has assigned an art adviser to make the planning. The art adviser first gathers the requirements for a *public-art* piece such as the amount of space, the preferences of the client, the purpose of the building, the theme of the art and the environment. The art adviser then searches on *public-art* artists and compares their portfolios containing examples of their work. The next step is to match the collected requirements with the artists. The art adviser needs to make a selection of several artists and then present these options to the client for approval: *"Which public-art artists match the project requirements?" [P4]*

2.4.3.3 Keeping up-to-date

We found keeping up-to-date is the type of non-goal oriented task carried out by the cultural heritage experts. There are two ways in which the experts keep up-to-date: by actively seeking for updates, or passively, using technology that automatically detects content updates and sends notifications.

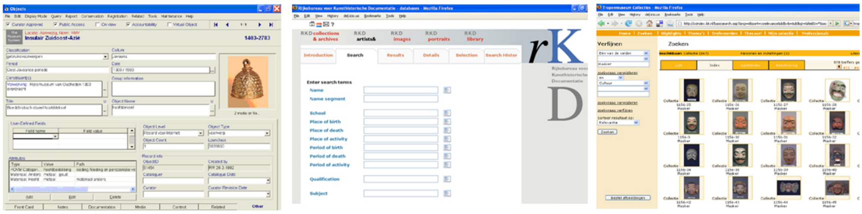


Figure 2.3: Interface screen shots from different types of sources, giving an impression of the typical complexity of the search interface: TMS collection management system (left), RKD online library (middle), Tropenmuseum's public website (right).

1. Active: Going to the sources and scan through for changes from sources (e.g. browsing). Experts keep up with the latest news on artworks of their interest, follow auction news, keep up-to-date with the price of artworks or with changing artwork ownership.
2. Passive: Using technology to automatically deliver new information from sources (e.g. RSS feed, email). Many experts subscribe to community mailing-lists to receive information on developments such as new exhibition announcements or reviews of new books and other publications.

2.5 Discussion

The results in the previous section allowed us to identify the experts' most important information seeking tasks. In this section, we identify a number of problems with using current tools, categorized by information seeking task. Furthermore, from our observations, we distill general trends of the information seeking needs in the cultural heritage domain that may extend to other domains.

2.5.1 Fact Finding

Experts' search questions can be simple or complex, with many constraints. Most search applications they use support both simple and complex search. The two most frequent Fact Finding problems are where simple search does keyword matching across all descriptions and returns too many results; and where advanced search specifies values as constraints and retrieves too few or no results. Difficulties in building queries can occur if the expert is not familiar with the correct controlled vocabularies from the thesauri that are used to describe the objects.

2.5.2 Information Gathering

More than half (63%) of the use cases we found can be classified as Information Gathering tasks. Many of these use cases are relatively complex when compared to Fact Finding use cases, and they are often related to activities such as preparing new publications, designing exhibitions or managing collection documentation (see table 2.4). Information Gathering contains very distinct use cases, which can be further classified as sub-tasks such as *comparison, relationship, topic, combination* and *exploration* search tasks. Topic search is the main type of search for experts that need to prepare exhibitions, write publications or document objects. In Information Gathering tasks, experts typically search with several different sources. They are forced to manually collect, examine and synthesize relevant pieces of information, because these higher level activities are not supported by their tools. For each Information Gathering sub-task, we give examples that illustrate this mismatch and describe how experts compensate for the lack of tool support:

1. Comparison. Current tools are often not geared towards comparing two objects side by side, and comparing sets of objects is an even harder task: *"For our exhibition, what objects from Aceh*[12] *that are missing in our own collection can we borrow from that museum?" [P3]*

When a curator would like to compare parts of her collection with that of another museum, she might prefer to pick up the phone and discuss the issue with the curator of the other museum directly. Curators report this is often more effective than trying to browse the other museum's collection website, especially when both curators know their collection by heart. However, relying solely on a curator's memory may not be wise for large collections: As one curator explains after the interview: *"For my own collection, there are around ten thousand objects, it is still possible to remember my own stuff, but I cannot imagine a curator remembering every detail if he has to take care of hundreds of thousands of objects." [P15]*

In such cases, experts are forced to look up the relevant objects themselves, and to do all the higher level comparison related tasks manually as these are not supported by their tools.

2. Relationships Search. In the example of the Rembrandt exhibition, our expert executed a comprehensive search for relationships between Rembrandt and other people, to create a story for the exhibition. In such cases, an expert typically consults many sources, such as art books, history books, birth records and

[12] Aceh is a region in Sumatra, Indonesia.

biographies. She takes notes of different names from one book, and some of these names may refer to the same person that are mentioned under another name in another source. She then has to cross reference to see if there is more information about the related persons found in other sources. *"I check old archives, history books, collect the names and make the connections." [P8]*
Finding such relationships requires a lot of time and energy. Sometimes the relationship is direct and easily found within one source. More often, however, the relationship is indirect, requiring searching through many sources and making interpretations along the way. These tasks have to be done manually because they are not supported by the expert's tools.
3. Topic Search. When an expert needs to find out everything s/he can about an individual object, e.g. when it is to be added to the collection, many online and offline sources need to be consulted. There is no direct way to obtain all the answers. Most search interfaces provide only keyword search, and rarely allow users to browse by topic or obtain suggestions for related results. One participant wanted to know, *"Are there any objects in the museum related to the African trade in the 17th century theme? [P17]"*. She tried many related keywords in combinations that she thinks may be included in the description of the object: *'trading africa', 'goudkust' (name of area), 'handel' (trade), '1799', 'west-africa', 'akan' (name of people who lived in the area), 'elmina' (name of a fort), 'weight', 'boeien' (chain).* Eventually she found a few carved metal weights that the traders used with a balance. She was quite sure that there should be many more objects about that topic but she could not find them. *"There should be chains and special boxes with ornaments (to put gifts in), but I couldn't find them."[P17]*
For topic search, simple and advanced search interfaces (see figure 2.3) are not sufficient. Users have to (almost randomly) guess related keywords, which is unlikely to lead to finding all desired answers.
4. Exploration. When experts need to do exploratory search, they rely heavily on their domain expertise. Similar to topic search, users will try out different terms, but only because this is what the search interface supports. The difference is that they try to be more exploratory with the chosen terms in the hope that they will obtain results serendipitously. Currently, experts are forced to rely on their cognitive skills and creativity to bring all this knowledge together. In the "Staircase" example, the expert tried different terms that she thinks will lead to promising results (e.g. *staircase art, landscape art, city planning art*). These terms are usually very specific, relying heavily on the expert's creativity, and the procedure itself is often based on trial and error and may lead to no result. Exploratory search is helpful when the expert is looking for new ideas, e.g. for an exhibition. The problem is that, while some tools may support browsing from one link to another, none supports the exploratory search task directly.

5. Combination. Finding a match between two pieces of information is challenging. For example, an expert needs to find an artist, who best matches the project requirements, from a list of candidates. She selected several artists who were potential candidates, collected samples of their work and presented her findings to the client. *"I select around 5 best artists who I think are suitable for the job, then I collect and present their portfolio to the client"[P4].* Much of the work involved in the combination task is done manually and relies heavily on the expert's experience and personal judgment. The problem is that this task requires diligence and takes large amounts of time and effort.

2.5.3 Keeping Up-to-date

Even though our participants did not mention this activity as a priority, they would like to be kept informed of cultural heritage news that is interesting for them, such as new exhibitions, new publications, social and professional events. Passive KUTD (e.g. being subscribed to a mailing-list or RSS feed) takes less effort than active KUTD (e.g. browsing websites) because experts receive notification only when there are changes. Our study suggests, however, that the usage of passive KUTD is low. The reason behind this is because not all systems provide support for passive KUTD and experts are not used to passive KUTD – this was mentioned only twice out of 110 use cases. Despite its utility, passive KUTD can be irritating, especially when changes are too frequent or insufficiently important, resulting in information overload. *"Well, I am subscribed to a mailing-list but I do not have time to read all those emails (laugh)."[P5].* As a result, only few experts feel the need to subscribe to mailing lists or RSS feeds. The problem is in finding a balance between providing the experts with the most recent and relevant information while not letting them feel overwhelmed by too much information.

2.5.4 General Trends

Even though we focus our study on the cultural heritage domain, we found several issues that indicate general trends in information seeking behavior that may be relevant and apply to other domains, such as humanity, social sciences and professionals (Bruce et al. 2004; Rosner et al. 2007; Tibbo 2002; Warwick et al. 2005).

Information Seeking Tools — We observe that the current tools used by experts do not fully support fact finding, information gathering and keeping up-to-date information seeking needs. We identify several reasons why this is: (a) difficulty in making complex queries (b) pieces of information need to be collected from many different sources before being synthesized by experts (c) data is mostly unstructured, making the retrieval process challenging.

Fact-finding requests require better data descriptions and tools that assist users in expressing complex queries. For most of the information gathering tasks, we feel current tools are insufficient because the tasks often rely on finding higher level relationships between individual facts, which are distributed across heterogeneous sources. Typical examples are tasks that require a combined view on the collections of two separate museums or tasks that combine information from the museum's collection management system with more general art-historic background information. In these cases, tools that can make the higher level relationships explicit are needed. Similar findings were reported by (Warwick et al. 2005), for humanity researchers where the search tools available are not adequate for their complex information seeking needs.

Searching Multiple Sources — In most use cases, experts need to consult multiple sources to obtain their answers (see Table 2.3). This is consistent with research found in other domains such as in (Rosner et al. 2007; Tibbo 2002). It is rarely the case that experts rely on only one source of information. For example, in Information Gathering tasks, experts need to constantly compare, relate, combine and explore information from different sources. Even though none of the experts complained about the tedious way of searching, we observe that they spend large amounts of time and effort on repetitive searching because they need to repeat the same query in different sources.

Communication — While experts consult multiple online and offline sources, personal contact between experts remains an important means of obtaining information (12 out of 14 information exchange use cases). We identify a number of reasons for this. Experts find it more convenient to contact other experts rather than to do the searching themselves; other experts have knowledge and experience that cannot be found in any document or information source; experts need to consult each other to find consensus on certain matters or decisions.
"We have regular meetings with other museums to decide which words should be in the thesaurus." [P2]. When asked about how she planned her exhibition: *"...I talk to many people who are experts in Rembrandt." [P6].*

Provenance and Trust — Another important issue mentioned by our experts in this context is source credibility: experts only mine information from sources they trust. This is also the case in other domains (Rosner et al. 2007). Generally these trusted sources are agreed upon within a specific community. *"For the thesaurus we decided to use all the literature which we agreed upon. Several years* *ago we did not accept anything from Internet, but nowadays some. The curators are a little bit afraid of it, who says the information is correct?"*[P2]. When systems use information from multiple sources, source credibility measure needs to be taken into account at both a functional and interface level to help users decide on the quality of the search results.

Information Maintenance — While our study focuses on understanding search behavior, we found that information seeking behavior is closely related to information maintenance. Once an expert finds a piece of information, she may want to store it because it may be reusable in the future. While cultural heritage experts adopt many strategies commonly used also by other professionals (Bruce et al. 2004), such as making bookmarks and keeping notes, there is no integrated way to store and maintain retrieved information from different sources that enables experts to re-access information efficiently. Thus, if we want to support these experts' information seeking tasks, we also need to help these experts maintain the information once discovered.

2.6 Design Implications

The results of our study suggest that cultural heritage experts need better support for complex query formulation, information gathering tasks, keeping up-to-date and for searching across multiple sources. In this section, we discuss potential research directions and the impact of this study on the design of the search tools developed within the MultimediaN project.

Information Seeking Tools
Fact Finding — Our results show that experts would benefit from explicit support for formulating complex queries. A *facet browser* (Hildebrand et al. 2006; Yee et al. 2003) is an example of an interface that can assist users in building complex queries in an incremental manner. The interface shows intermediate results as the user builds a complex query step by step. Yee et al. showed that facet interfaces work well for visual resources described by structured data (Yee et al. 2003). Cultural heritage information is often visual (e.g. photographs of museum objects, visual archives and photos of artists) and stored in structured museum databases. Deployment of a facet browser interface, however, typically requires extensive configuration of data and software. This requirement makes facet browsing less compelling for museums, which often use more generic museum management systems. Facet browsers also may not work well on heterogeneous

datasets because the number of facets tends to become too large. */facet*[13] is a facet browser that we developed for heterogeneous domains (Hildebrand et al. 2006). We are currently researching how we can effectively use facet browsing within cultural heritage organizations.

Information Gathering — Due to the complexity and diversity of information gathering search activities, different specialized search tools may be required. For example, to support experts searching for relations among artefacts or concepts, the sources used need to be structured and linked to each other, providing relationships among museum entities, such as between artists, art styles and artworks. Our *Relation Search*[14] prototype allows users to find connections between any two objects or concepts, such as between an artist and an artwork or between two artists. The results are organized from the shortest (showing direct relations) to the longest (showing indirect relations). Current challenges in this type of search is on how to differentiate the "interesting" from the large amount of "trivial" relations, since these notions are subjective and context dependent.

Another example is topic search. Whenever experts search for information centered around a particular topic, they need information related to a single term as well as suggestions for related terms, e.g. nearby geographical locations or related cultures. While recommender systems based on content-based and collaborative filtering have already been developed for suggesting relevant research papers (Torres et al. 2004; Schafer et al. 2007), further research is needed to explore if and how such approaches can deployed in the domain of cultural heritage.

Keeping Up-to-date — While experts need to keep up-to-date with the development of particular topics, they should not be overwhelmed with irrelevant information. User interfaces developed for this purpose, such as RSS readers, already exist. Our study, however, shows that only few of our experts have actually used them. This may be because RSS readers are relatively new, and over time more domain experts may start to use them. One thing that may be helpful for subscribers is a tool to manage the frequency of update and provide filtering and/or ranking based on, e.g., priority and topic. Further research is needed to explore to what extent keeping up-to-date functionality can and should be integrated into tools that experts already use.

Searching Multiple Sources

For many experts, a "meta-search" functionality that simply aggregates search results from multiple sources would already be a great improvement over the

[13] `http://e-culture.multimedian.nl/demo/facet`

[14] `http://e-culture.multimedian.nl/demo/path`

current practice of repeating the same query in isolated search applications. Our project's demonstrator (Schreiber et al. 2006) allows users to search and navigate multiple collections from multiple organizations. In addition, our system uses available domain background knowledge to find semantic relationships between items from different collections. We are currently researching how to effectively explore these relationships in both the search algorithm and in the presentation of the results.

Provenance and Trust
Trust in the information source is a key aspect for experts, so applications providing access to multiple sources need to explicitly communicate the origin of all information to the user. How provenance can best be conveyed when information from many sources needs to be displayed is still an open research question. As a first step, we conducted a study on how visualization of the credibility of the cultural heritage sources influences peoples' decisions (Zhang 2007). This study showed, for example, that visualization of source credibility ratings significantly increased the confidence of users that were presented with information from multiple sources. Additionally, trust measures could also play a role in the search functionality itself, for example by ranking results based on the credibility of their source. Further research is needed on the design and evaluation of this type of functionality.

2.7 Conclusions and Future Work

We have presented the results of our study on the information seeking needs of cultural heritage experts. The goal of the study is to understand: why do cultural heritage experts search; where do they search for information; and what are their information seeking tasks. Our study suggests that experts' daily search tasks are dominated by (high level) Information Gathering while the search systems they use support (low level) Fact Finding tasks. As a result, experts need to compare, relate and combine pieces of information manually or ask their colleagues. We also found that while the experts have simple as well as complex questions, their current tools provide insufficient interface support for query formulation. In addition, most experts' search tasks require information from many different sources, while their tools tend to support search in only one source at a time. Finally, we discussed the study's implications on the design of the search tools developed in our project.

Our next step is to concentrate on each of the tasks and see how to improve the search experience. We will focus our attention mostly on Information Gathering, since this is our experts' main search activity but seems to have the least support.

We will investigate the different ways to present cultural heritage information that enable our experts to compare, relate, explore and combine information and search for related topics. Furthermore, we plan to perform user evaluations on our proposed interfaces to verify that we have indeed helped our experts in their search.

CHAPTER 3

Lay Users' Information Needs: A User Study on Location-based Mobile Search

We investigate the information seeking tasks of lay users. We focus on location-based information seeking tasks using a mobile search engine. The aim of the study is to capture lay users' location-based information seeking tasks, the information sources that they use and the social and temporal context that influences their search (research question 1). To capture users' location-based search tasks in a ubiquitous setting, we used a longitudinal ethnography study combining log analysis with diary studies and interviews. This study shows that lay users' information seeking tasks consist of fact finding, information gathering and non-goal-oriented tasks. We identify the different information sources lay users use for location-based mobile searches, the challenges they face when carrying out these tasks and the factors that influence their search decision process (research question 2). On the basis of these, we propose suggestions of features to enhance location-based services, such as location-based query refinement, location-based service mashups, and iterative and comparative support tools.

This chapter was the result of an internship work at Google.com and was published as "Fancy a drink in Canary Wharf?": a user study on location-based mobile search" in the 12th IFIP TC13 Conference on Human-Computer Interaction conference INTERACT'09 (Amin et al. 2009), and was co-authored by Sian Townsend, Jacco van Ossenbruggen, and Lynda Hardman.

3.1 Introduction

While the World Wide Web allows access to information globally, local geographical aspects are nonetheless important in many web search tasks. In a generic search, such as searching for a movie trailer or a book, geographical aspects are irrelevant. This is not the case for location-based searches. For example, when a user is searching for the nearest pubs, the system is required to identify local pubs and only present those that are in the neighborhood. Unlike generic search, the geographical context is important for location-based search. Much research has shown consistent demand for location-based information on the desktop as well as on mobile devices (Kamvar and Baluja 2006; Spink et al. 2002; Yi et al. 2008). Estimates on how often such search occurs have been disclosed by several major search engines: in (Spink et al. 2002), samples of queries from 2001 Excite desktop searches were examined and 19.7% of them were searching for places, people and things; around 9-10% of the queries collected in (Yi et al. 2008) on Yahoo! mobile search were identified to have geographical search intentions, whereas more than 15% of 1 million Google queries from PDA devices are for local services (Kamvar and Baluja 2006). In addition to search engines, many web services offer search support for diverse location-based information, such as, local businesses search and review (e.g. `www.yelp.com`), city guide (e.g. `www.citysearch.com`), or local traffic news (e.g. `www.highways.gov.uk/traffic`). These examples show a healthy market for location-based information providers. Research on location-based information needs is still at an early stage and most reports are confined to the different domains of interest related to location-based search. For example, large scale mobile query analysis done by Google (Kamvar and Baluja 2006), Yahoo! (Baeza-Yates et al. 2007) and an EU Mobile operator (Church et al. 2007) consistently report that people make location-based queries on a wide range of domains of interests, such as food and drink (e.g. restaurants), shopping (e.g. stores) and travel (e.g. addresses). A small scale diary study on mobile information needs (Sohn et al. 2008) reported that people search for generic information, such as looking for music, and for location-based information, such as looking for places of interest (POI), business hours, and movie times. In these types of searches it might be useful to consider location context as it has proven to improve the quality of query prediction in a mobile search application (Kamvar and Baluja 2007). Unfortunately, past research on has not given any explanation as to *when, where* and *why* people conduct location-based search. Most research on mobile information needs do not focus on this topic e.g (Baeza-Yates et al. 2007; Church et al. 2007; Kamvar and Baluja 2006; Sohn et al. 2008), and those that do, limit their approach to automatic query log analysis e.g. (Sanderson and Kohler 2004; Sanderson and Han 2007). As far as we know, there is no user study that exclusively investigates peoples location-based search motivations and context. In order to improve and optimize location-based services, it is necessary to understand peoples location-based information needs and the context in which they occur. We define location-based search as: *"Search for a business or place of interest that is tied to a specific geographical location."* This definition is somewhat broader than Location-aware search where it is implied that the system has knowledge of and exploits the searcher's location. We use this terminology to reflect the type of query that is collected from the search engine in this study. In this research, we investigate location-based search on mobile devices. The contribution of this chapter is an in-depth investigation of location-based search behavior using the mobile search engine and includes the spatial, temporal, and social contexts in which this search occurs. We look into how the users location and the people they were with influence the location-based search made on a mobile device. We collect comprehensive information from search engine log data, location data tracking and diary entries. Key findings include, first, that people specify location-based search at different levels of granularity, from simple to detailed queries, constructed by different types of information. Second, people tend to travel along regular routes in their environment and visit the same places of interest regularly, and the impacts on their search behavior. Third, most location-based searches on mobile devices are conducted in the company of other people, such as friends, family or colleagues.

This chapter is structured as follows. In the next section, we describe our research method. We then present the results, including how people express location-based searches, the contexts in which the searches are conducted, and the search tools used when they do location-based search. This is followed by discussions on the key findings and design considerations to improve location-based services. We close with a summary and future work.

3.2 Research method

We want to investigate how people express location-based information needs on a mobile device, what are the different situations when this search occurs and what are the search tools that people use to conduct location-based search. Our research questions are as follows: What types of location-based search can be identified? In what context (spatial, temporal, and social) are these searches initiated? What are the information sources (e.g. maps) used for location-based searches? Capturing users' behavior on a mobile platform is challenging because of the difficulty of unobtrusively collecting data in a ubiquitous environment. Previous research with the mobile phone relies on different methods, such as interview e.g. (Cui and Roto 2008), log data analysis e.g. (Baeza-Yates et al. 2007), (video) observation e.g. (Mark et al. 2001), experience sampling e.g. (Consolvo and Walker 2003), diary study e.g. (Brandt et al. 2007; Sohn et al. 2008), or a combination of two or more methods (Riegelsberger and Nakhimovsky 2008). Search logs provide data on users' realistic search engine usage. They do not, however, provide any insight into users' intentions. Identifying location-based information needs solely from log analysis is not easy because the intention might not always be expressed explicitly in the query by specifying an address or a city name (Yi et al. 2008). Diary studies give the opportunity for users to express search goals. Our approach is to combine two methods: search logging and diary study in order to benefit from the strengths of both. Additionally, to understand how people perceive the spatial world around them and the places that are important to them, we use a method similar to that proposed in (Lynch 1960) in the form of a creative exercise where people describe and visualize their environment and the places they go. Next, we describe the web diary tool, the user study procedure and participant profile. The Web-based Diary Tool The diary tool collects and links 3 types of data: users search event logs from the Google mobile search engine, location tracking data from participants device and diary entry data.

Search event log. Events occurring using Google mobile search[1] were collected (e.g., queries entered, clicks, scrolls, keystrokes) including the corresponding timestamps. We also collected SMS snippets. Participants were asked to send an SMS to a dedicated number whenever they identified an information need that was answered by some other information source (e.g. maps, other websites) or a need that could not be answered.

Location tracking. We logged the users location (latitude and longitude), every time a search is made. The location data provides information on the participants whereabouts when the search occurs.

Diary entry. Participants were required to log into the diary at the end of every day and to answer questions about their search activity. The participant's detailed search history throughout the day is made available. There are 2 steps of action in the diary tool. First, participants need to identify search tasks. In some cases, information needs can only be answered by a participant through conducting several queries in a search session. For example, a user might type several queries, such as *pubs, bars, Irish pubs*, for one search task: to find the nearest pubs. Thus, in the web-diary tool, we provide an option for the participant to group these queries from the same session into a single search task. Second, for every search task,

[1]Software was installed on users phones to enable us to log search events and location (latitude and longitude) while users were searching with the Google Mobile Search `http://www.google.com/mobile/`

participants need to answer several questions:
How important is this task?
Where were you?
What were you doing?
Who were you with?
Describe in detail what you needed to look for.
When was this need initiated?
What tool(s) did you use to find your answer?
Did you successfully accomplish your task?
Positive and negative experiences with the tool used?
The search event log provides information on the location-based information needs that occurred and the search history. The location tracking indicates where the search took place, and the answers provided by the participants from these questions give an overview on the condition and situation when the information need arose.

3.2.1 Procedure

The study is divided into four parts:
Pre-study interview. Participants were briefed about the procedure and were requested to provide background information about their daily activities and search experiences. Participants were free to search for any topic but were told that this study was investigating location-based search. Participants gave their consent to have their search activity logged and location tracking during the course of the study.
Digital diary study. The diary study lasted for 12 days. The duration of the study covered different types of days (6 work days, 6 days weekend/public holidays). During this period, participants were asked to make approximately 3 queries per day, as a loose guideline, with the search client on their phone, and to log into the diary website daily from their PC to answer questions related to the mobile searches made.
SMS Reminder. During the course of the diary study, an SMS was sent every 2 or 3 days as a reminder to make queries, and to log into the online diary daily.
Post-study interview. After the study, participants were invited for a semi-structured interview. They were asked additional questions relating to their mobile search experiences based on the answers provided on the diary website. Any unclear or missing diary entries were clarified. Finally, participants were given a creative exercise to sketch typical places where they usually go, and to specify the distance and means of travel to get there.

3.2.2 Participants

Twelve participants were recruited from the London area from which 9 participants (5 female, 4 male) successfully completed the whole study. We recruited participants who used Blackberry phones[2]. All nine participants were young professionals, aged between 27-35 years old (M=32.6, SD=2.5). We recruited participants with different occupations, such as a financial advisor, photographer, and retail sales person. All participants were regular users of mobile phones and PCs. Our participants use their phones for calls and SMS-es on a regular basis. In addition, they have a mobile Internet subscription and use it to access their emails, news and entertainment content (e.g. BBC), to plan journeys (e.g. maps, public transport schedules), to keep in touch (e.g. Facebook) and to do search (e.g. Google, Yahoo!).

3.3 Results

During the course of the study, we collected 347 location-based mobile search queries (see Table 3.1). These were organized by participants into 186 location-based search tasks (see Fig. 3.1b). Thus, in average, participants make 1.87 queries per location-based search task. Additionally, there were 13 search tasks (6.3%) which were not location-based search tasks, such as downloading images, music, or games (i.e. Transaction task[3]). We omit these and focus our analysis and discussion on location-based search tasks. The diary tool provides a rich context (who, where, when, why, and how) for each search task conducted. Queries and tasks were systematically analyzed as follows. First, we examine the queries from different perspectives to understand how people express location-based information needs, the different domains of interest and different types of search tasks. Second, we examine the answers provided from each diary entry to understand the spatial, temporal, and social context when location-based information needs arose. Third, we examine the tools that people use to find location-based information and the reasons behind unsuccessful location-based search. Finally, we use affinity diagram technique to understand emerging topics that lead to the discussions and design implications.

[2] Due to a technical restriction of our system

[3] Transaction task is an information exchange task, e.g., online banking, downloading multimedia documents. Transactions are often associated with a user name and password combination (Kellar et al. 2006)

3.3.1 Types of location-based queries

The queries were analyzed from 3 different perspectives. Each perspective offers a unique view on location-based search. We examined the query patterns to identify how people express location-based information needs (Syntactic view). Afterwards, we looked at the variety of domains to get an overview of peoples' interests and on the types of places people look for the most (Domain view). Finally, we look from the perspective of information seeking tasks, to understand the breadth of the search goal (Task view).

3.3.1.1 Syntactic view

We manually analyzed each query to investigate how people express location-based queries (see Table 3.1). There are 6 different components of a location-based query: business name (e.g. Orange Tree pub), business category (e.g. Mexican restaurant), location name (e.g. Pennyhill Park), event name (e.g. Cbeebies tour), product or service (e.g. Smart 12 month lease), and web address (e.g. `www.streetmap.co.uk`). We found that participants express location-based information needs in different levels of granularity with these elements: (a) *A simple location-based query* consists of only one of these components (see Table 3.1, a-e). Additionally, some people make Navigational queries[4] to search for a website that contains location-based information (see Table 3.1, f-h), e.g. `www.streetmap.co.uk,www.tfl.gov.uk/tube,` or `www.kidslovelondon.com`. These sites usually offer dynamic and detailed information, such as tube schedules, the city weather forecast, and local events. (b) *A detailed location-based query* consists of a combination of more than one element, e.g. business name and location name (see Table 3.1, i-n). From Table 3.1, we observe that location-based information need is expressed by specifying a business name (22.9%), a business name and location (13.0%), or a web address containing local information (12.1%).

3.3.1.2 Domain view

The queries were further analyzed to identify the domains of interest. Four main domains of interest that our participants looked for (see Fig. 3.1A): stores (27.0%): businesses that offers products/services, such as electronic stores, furniture stores, book stores; food & drink (24.5%): businesses such as restaurants, pubs, cafes; entertainment (13.7%): such as cinemas, theaters, concerts; and transportation (10.3%): public or private companies, such as train, bus, tube, taxi.

[4]Navigational query is a query where the intention is to reach a particular site (Broder 2002).

Table 3.1: Local query construction (Total: 347 search queries)

	Query	Examples
67.7%	**S**imple location-based query (235 queries)	
22.9%	a.business	*the orange tree pub*[P4]
4.3%	b.business category	*mexican restaurant*[P6]
2.4%	c.event	*cbeebies+tour*[P5]
7.2%	d.product/service	*smart 12 month lease*[P6], *vintage leather jacket*[P7]
2.9%	e.location	*where is covent garden[P3], pennyhill park[P4]*
10.1%	f.local news, weather	*BBC weather*[P1]
5.8%	h.transport schedule, map	*train times*[P1], *tube map*[P4]
12.1%	h.url	*streetmap.co.uk*[P4], *kidslovelondon.com*[P11]
32.3%	**D**etailed location-based query (112 queries)	
1.4%	i.business & attribute	*tesco opening hours*[P1]
13.0%	j.business & location	*burger king charing cross*[P6], *nandos - se19*[P3]
5.8%	k.business category & location	*canary wharf bars*[P2], *model agencies in london* [P7]
1.9%	l.event & location	*moon walk 2008 london*[P7]
6.8%	m.product/service & location	*driving testing centre-hither green*[P3]
3.4%	n.multiple locations	*directions from green park to primrose hill* [P7]

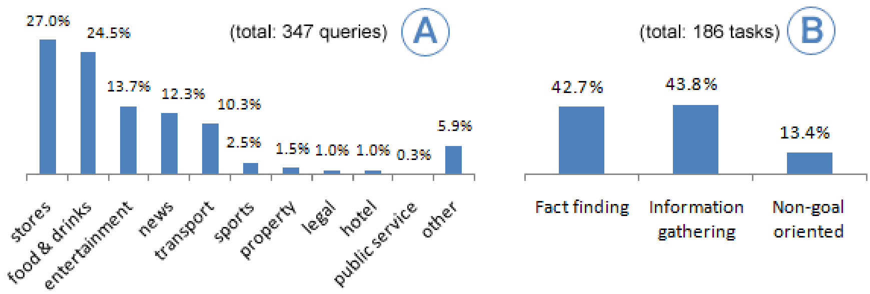


Figure 3.1: a) Distribution of interest. b) Types of location-based search task.

3.3.1.3 Task view

The diary entries provide rich descriptions of users' search tasks and motivation behind each search query. We found there are different location-based search tasks (see Fig. 3.1b) from specific to broad as explained by the Information Seeking Task

taxonomy (Amin et al. 2008; Kellar et al. 2005; Kellar et al. 2006). These tasks are:
(a) Fact finding tasks. The task is goal oriented and focused: Users look for specific factual pieces of information. The search intention is usually straightforward e.g. looking up a phone number. We found 42.7% of the location-based search tasks fall into this category. Examples of participant's tasks are: Looking for service information: public transport timetable. *"I want to know the train times home to return to London"* [P1]. Looking for business information e.g. contact details. *"I needed the number for Pizza Express to order the food."* [P12]
(b) Information gathering tasks. Users carry out several search tasks to fulfill a higher level goal. We found 43.8% of the search tasks belong in this category. Examples of participants tasks are: making a decision where to go for dinner by exploring different businesses with different constraints, such as cost, distance and product. *"Looking for a Mexican restaurant for a surprise birthday party, found a good selection of restaurants with distance information and maps to easily locate."* [P6] . *"We were looking for a reasonably priced place to go and eat in central London as that would be convenient for everyone."* [P11] From our search logs, we observe that Information gathering search tasks tend to be iterative (the user goes through several rounds of searches), *exploratory* (the user tries out different alternative queries/search strategies), and *comparative* (comparing different search results before making a decision).
(c) Non-goal oriented information seeking tasks. Users' information needs are not goal driven. The motivation of the search activity is either to be informed or to see what is new or interesting. We found 13.4% search tasks that belong to this category, e.g. Keeping up-to-date with upcoming events. Users carry out serendipitous browsing to see if there are interesting events. *"I was just trying to find out if there were any packages for dinner and show, or if there were any specials in restaurants, bars, clubs.. just to get an idea and see if I find something that stands out."* [P7]

3.3.2 The context of location-based search

The web-based diary tool gave us comprehensive information on where, when and how each location-based search occurs. We discuss each of these in the following subsections: spatial, temporal and social context.

3.3.2.1 Spatial context

Participants' search activities and locations, while using the search engine, were logged throughout the study. We found that many searches are made either at home, at work, while travelling between the two, or at regularly visited places. With the exception of a weekend break, most of the time, participants' follow regularly used routes and regularly visited places (hotspots). Fig. 3.2 (lower half) shows a map visualization of searches made by a participant. Searches that have overlapping locations are represented by one point. The line connecting home (1) and work (2) represents a participants daily travel routine with the train. At the post-study interview, we asked participants to make a sketch of the places they most frequently visit. For example, Fig. 3.2 (upper half) shows the corresponding sketches from the same participant. His regularly visited places are home, work, and several favourite businesses (pubs, restaurants, cinema, cafe) that are located within 5-15 minutes walking distance from home or work. By comparing the map visualization and the user sketches, we have a better understanding about our participants spatial mobility, their hotspots, the distance between hotspots and their means of transport. From the 186 search tasks made by participants, the most common place to search was: at family/friends' home (6.5%); public places (8.5%), e.g. at the gym, at the pub/cafe; at work (12%); on the move (20%), e.g. on the public transport, inside a car; at home (53%), e.g. watching TV, lying in bed, preparing to go out. We discovered that the target location is more often related to their regularly visited places (e.g. work, home) rather than to the proximity of their current location, e.g. while on a way to work a participant searched for an optician close to home/work where he can easily stop by: *"I needed to find the closest opticians and deals nearest to home or work."* [P3].

3.3.2.2 Temporal context

For every search task, participants were asked to specify whether the need for the information was spontaneous, related to something planned on that day, or arose in relation to something prior to that day. We found that more than half of the queries were a spontaneous need (66.1% of 186 tasks), prompted mainly by activities and situations. They were based on recently acquired needs to obtain information and required immediate answers. *"I needed to find a phone number for a client to contact him urgently."* [P5] . Less than a quarter of the queries were needs that were planned for the same day but had less sense of urgency (21.5% of 186 tasks), e.g. deciding where to go for lunch later. *"Wanted to book a table for lunch with a local Tapas restaurant."* [P1]. The remaining search tasks were needs for another day (12.4% of 186 tasks). These tasks are mainly exploratory search and not urgent at all. *"I was looking at a health farm for a weekend break."* [P4]

Table 3.2: Decision making in local search (Total: 76 search tasks)

	Reason
23.7%	has a particular product/service
15.8%	is recommended by other people
13.2%	is decided together*
7.9%	is close to where I am now
7.9%	it's my favorite place
6.6%	is close to my home
6.6%	is decided for work purpose
5.3%	is decided by someone else
3.9%	has the best offer (e.g. sale)
9.2%	other (e.g. serendipity search)

* e.g. with friends, family, colleagues

3.3.2.3 Social context

Participants were asked about who they were with at the moment they conducted the searches. More than three quarters of the location-based search tasks were conducted in the presence of others (76.1% of 186 tasks), e.g. while with family, friends or colleagues. Most location-based searches were prompted by: conversations with people, *"We were talking about Rosarito and we wanted to know where that is."* [P7]; event planning, *"We were looking for somewhere to go for lunch on Easter Monday."* [P5]; recommendation by people, *"We have spoken about this wine tasting place and I wanted to go and have the experience."* [P11]; and necessity, *"We needed directions on how to get there."* [P11]. Less than a quarter of location-based search tasks were conducted alone (23.9% of 186 tasks), mostly driven by necessity. *"I wanted to see that days weather."*[P5], *"I wanted to know how to get from Hyde Park Corner to the City Airport."* [P7]. Because of the common perception that a mobile device is a personal device, we were surprised to find most location-based searches are not merely a solitary activity but one that is strongly influenced and triggered by social interactions. Cui et al. (Cui and Roto 2008) reported similar observation: that the mobile web acts as a *conversation enhancer*, such as to start a new topic, an ongoing discussion or settle a dispute. In our study, the mobile device plays a bigger role than merely as a communication tool and a conversation enhancer. Mobile search supports social activity, such as searching for ideas, collaborating on making plans, and sharing recommendations.

Related to the influence of spatial, temporal and social context, we investigated the decision making process in location-based search. In our post-study interview, 8-9 tasks were randomly chosen for every participant. For these tasks, participants were asked about the reason(s) behind choosing a certain business from the search result list. Table 3.2 shows the 9 most frequent reasons for choosing a certain place of interest. Most participants made the decision based on the availability of a specific product/service (23.7%), e.g. checking if the store sells a particular product. *"Needed to stock check a toy for my niece's 1st birthday."*[P1]. Many of the decision making processes in location-based searches were strongly influenced by social elements, such as recommended by other people (15.8%) *"Wanted to confirm the location of a pub. Had been given the name by a friend."* [P1], or decided together with other people (13.2%) e.g. *"It's a restaurant to go for our anniversary."* [P4].

3.4 Location-based Information Sources

Where participants were unable to satisfy their information need with the Google service, they were asked to report any other tools they used. A number of various location-based information sources were used, such as: business directory, e.g. `yell.com`; event guide website, e.g. `www.timeout.com`; travel website, e.g. `www.nationalrail.co.uk`; news website, e.g. `www.bbc.co.uk` and information in public places, e.g. poster or map on the wall. Participants were also asked to note down if the search was successful or not. Of the 186 search tasks, 64.1% were successful, 24.3% were partially successful, and 11.7% were unsuccessful. Several reasons why location-based search on a mobile device was unsuccessful are, e.g. (a) Participant could not find local content or could not find up-to-date information. *"Mainly American sites, couldn't find the UK one easily and then couldn't get the phone number."* [P5]. *"A very old wrestling information came up. Like years old."* [P11]. (b) Difficult to conduct complex search with mobile interface. *"National Rail Enquires needed the same information entering 2/3 times which was annoying and time consuming."* [P1]. (c) Problems with mobile connection *"It timed out a few times so I gave up."* [P4], *"Before I could finish my search, my train went into a tunnel and therefore no mobile connection."* [P6]. When the search is unsuccessful, the participant has the choice to either abandon the search altogether, look for another information source, or develop a new problem solving strategy. *"On Mobile, if I don't find something quickly then I give up, I lose interest. It would be rare to go back (to refine query). I guess it's cos you're used to going back using the back button on your PC ...on mobile you just give up."* [P6], *"There was nowhere I wanted to go in the search so I just decided to drive to the area and decide then."* [P4].

3.5 Discussion

Our goal was to investigate location-based search needs and to understand the context in which they occurred. From the results, we distill a number of key findings:

Most location-based search is conducted in the presence of others — Although a mobile phone is a personal device, our study shows that more than three quarters of the searches are done in the presence of other people (76.1% of 186 tasks). In some cases, these searches are *group information needs* rather than individual needs. This is because location-based mobile search is closely tied to social/group activities, such as going out. We also discovered that decision making in location-based search (e.g. where to go, which businesses to choose from) is highly influenced by social factors, such as recommended by people they know or decided together with family/friends.

Search now vs. search later — Our study reveals that mobile information seeking behavior progresses over time. A study on the mobile consumer behavior in 2001 reported that even though the information needs were triggered at a particular moment, people prefer to do information gathering search activities at a later time with their desktop, because of the difficulty in acquiring comprehensive information and making product comparison on the mobile phone (O'Hara and Perry 2001). In our study, this is partially still the case. Comparison search is still a difficult task to do on a mobile device. However, accessing information on the mobile device has become easier. On many occasions, participants prefer to use their mobile device at home rather than using their desktop because of the *lower engagement threshold*. For example, one participant and his daughter used the mobile search to plan their day while having breakfast in the kitchen. On another occasion he used mobile search just before bedtime while discussing weekend plans with his wife. These quick searches were considered inconvenient to do with the desktop computer located in the other room. Moreover, as they can now access many things on conveniently through the mobile device, there is less need to plan or to do search ahead of time. One participant puts it: *"It will change a lot of things, you usually do your homework before (at home), now you can do it afterwards (on the road)."*

Resident vs. Tourist — It should be noted that there are differences between the users in this study and the users in studies on mobile tourism, e.g. (Brown and Chalmers 2003). In our study, participants are residents, are somewhat familiar with the city, conduct location-based search regularly, and search on a broad domain of interest from plumbing to paint balling. Tourists, however, are more likely to be unfamiliar with the city (need to orient themselves often), have fewer time restrictions, and search for specific domains of interest e.g. city landmarks

and how to get to these places (Brown and Chalmers 2003). We expect differences between the two user roles with respect to location-based search types, and their spatial, temporal, and social context.
Study limitation — We acknowledge that there are several limitations to our study. First, our participants are Blackberry users. As pointed out in (Kamvar and Baluja 2006), users with Qwerty keyboards may have different search patterns compared to users with 12-keypad devices with regards to the distribution of number of words per query, domain of interest distribution, query distribution and session characteristics. Nevertheless, we expect users location-based information needs and the aspects investigated in this study (information seeking tasks, the role of spatial, temporal and social context) are mobile device independent. Second, our study had a relatively homogeneous demographic profile. This should not come as a surprise. The reality is that smart phone and mobile Web penetration is highest amongst a fairly narrow group who tend to be young professionals. It would not have been logical to recruit a cross section of all mobile users since we wanted to ensure we only recruited people who already use and were thus familiar with mobile search. It should be acknowledged that the findings of this research may vary for different user segment. Nevertheless, this research is among the first to report on location-based search behavior and actual usage.

3.6 Design implications

Location-based search experience can be improved by taking into account users' motivation, search patterns and context. We present a list of issues that should be addressed in future system designs.
Detecting and predicting location-based information needs — As mentioned in (Yi et al. 2008), it is difficult for a search engine to identify whether a query is location-based without any location qualifier. In our study, we found that most of the time participants only specify business name, business category, event name or product name. Thus, based on our findings we recommend that whenever a search engine receives a query that contains one of these elements, the search engine should prioritize using location-based information.
Recommendation based on hotspots — We found that people move along regularly used routes and their location-based search interests are within their hotspots (as shown in Fig. 3.2). Moreover, users' interests in location-based information are usually within the proximity of their hotspots. Thus, a personalized search result that provides location-based information tailored to users hotspots areas would be potentially valuable.
Location-based search query refinement — Features that help avoid making

mistakes when typing unfamiliar business/place names and addresses is useful to have in search engines. Search engines can help users by providing query suggestions based on likely local business/place names, for example, a list of query refinements, similar to the *'Did you Mean?'* feature for business/place names that are phonetically or morphologically similar to what has been typed. Another useful feature to help users specify business names and places is to provide autocompletion suggestions of businesses nearby or favorite businesses from users' bookmark or search history.

Support iterative, exploratory and comparative search activity — Our results show that a large portion of our search tasks are information gathering tasks (43.8%, see Fig. 3.1b). Most search interfaces, however, do not provide features that support iterative, exploratory, and comparative search activities. Features that would help users in this task are: first, support users to collect, filter, organize, compare, save and share location-based search results; second, support exploration by allowing users to tweak a set of constraints, such as enable to filter points of interest by distance, business category, service price, in order to find the optimum search results.

Location-based services mash-ups — Integration of different services, such as business directory (restaurant, pub, store), places (car parks, bus stops), public transport system (train, bus, tube), navigation system (route, distance), other services (ATM, gas station, restroom) and map application is another way to help users in information gathering tasks. Integration of different services will help them decide where to go, match their schedules, estimate distance and travel time between places, and ultimately help users make better plans.

Recommendation based on social network — Recommendations made by people play an important role in prompting location-based search needs and the decision making processes. It would be helpful to provide several features: (a) enable people to search, recommend and share experiences on businesses and make this information easily accessible to people from their social networks; (b) to provide location-based recommendations based on the interests of the whole group, for example if people within a social network are detected to be in close proximity to each other, a search engine can recommend places of interest that incorporate the interests of all members of the group rather than just the searcher.

3.7 Conclusion and future work

We conducted an in-depth web-based diary study using different types of data collected from search event logs, location tracking and diary entries. We found that location-based searches are mostly based upon just-in-time information needs that are usually related to social activity. Our study also shows that participants have regularly used routes and regularly visited places (hotspots). We also found that the target locations for these searches are more often related to users regularly visited places (e.g. work, home). Services that support location-based search need to take users social and spatial context into account. There is still much work to be done on how to properly implement search recommendations based on users' hotspots and social networks, e.g. with respect to information privacy and security. Finally, as mobile web searching becomes more widely adopted, we believe that location-based search will be an even more prominent need that calls for alternative mobile search interfaces and new engaging ways to interact with location-based information.

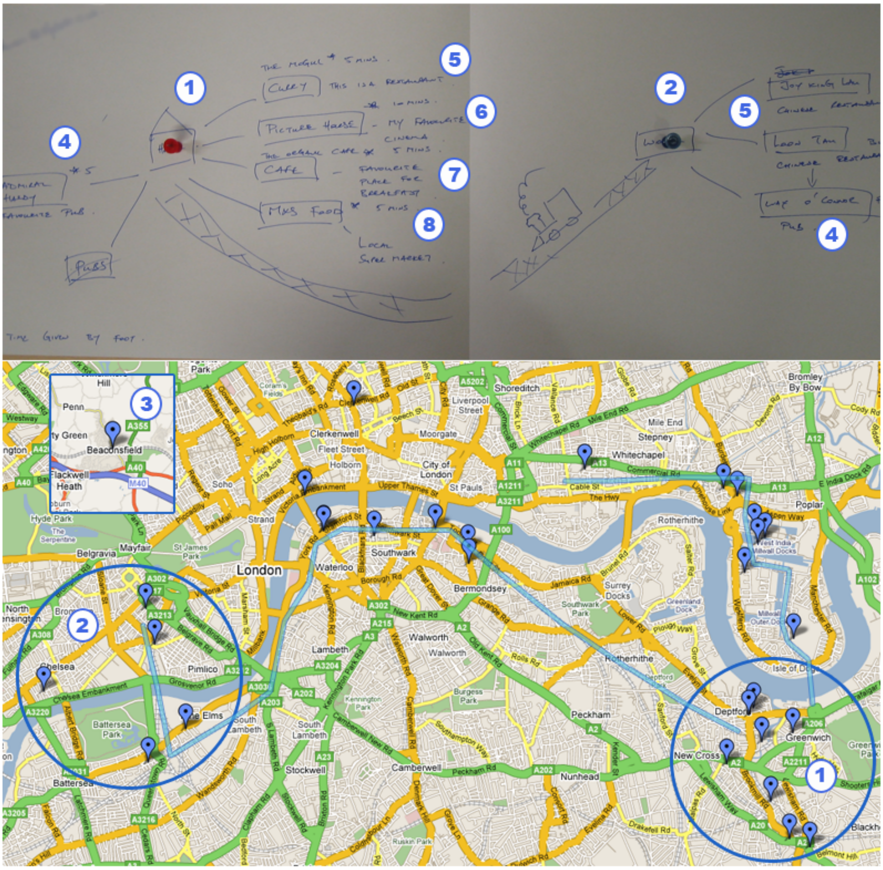


Figure 3.2: *Upper half:* An example of a participant's sketch about his regular routes. *Lower half:* the related location tracking data. The markers represent approximate search locations. Multiple queries from one location are presented as one marker: (1) home area, (2) work area, (3) weekend holiday out of town, (4) pubs, (5) curry restaurant, (6) picture house cinema, (7) cafe, (8) supermarket. The dashed lines connecting areas (1) and (2) are the participant's daily train route.

Part II

Supporting information seeking tasks across multiple sources

CHAPTER 4

Supporting Query Formulation with Thesaurus-based Autocompletion Interfaces

We investigate interfaces that assist a specific activity of information seeking, namely query formulation (research question 3). We carry out two user studies on thesaurus-based autocompletion. In the first study, we explore different strategies for grouping autocompletion suggestions using two different thesauri. The aim of the first study is to see which Group strategy is preferred and better understood by lay users. The results suggest that the best strategies depend on the different thesaurus structures. Group organization seems to be more appropriate for geographical names thesaurus (a domain specific thesaurus with symmetrical hierarchy), while Alphabetical works better for object names from a lexical database of the English language (a thesaurus with less symmetrical hierarchy covering a broad scope). In the follow-up study, we compare three different organization strategies (Alphabetical, Group and Composite) for location name search tasks. The aim of this study is to investigate the different organization strategies support users in query formulation. Based on this study, we derive guidelines to design thesaurus-based autocompletion interfaces (research question 2). The results of the second study indicate that for TGN thesaurus, Group and Composite organizations help users search faster, and are perceived easier to understand and to use than Alphabetical.

This chapter was published as “Organizing Suggestions in Thesaurus-based Autocompletion Interfaces” in the 31st European Conference on Information Retrieval (ECIR’09) conference (Amin et al. 2009) and was co-authored by Michiel Hildebrand, Jacco van Ossenbruggen and Lynda Hardman.

4.1 Introduction

There is a lot of interest in the Information Retrieval community for interactive query expansion features that can help improve user search performance and the quality of queries submitted. There are two types of interactive query expansion: real-time query expansion (provide suggestions during query formulation) and post-query expansion (provide suggestions after query formulation). Between the two types, real-time query expansion (RTQE), such as autocompletion has been most adopted in many operational search applications e.g. *Google Suggest* or *Yahoo! Search Assist*. RTQE is an attractive feature because it can reduce the required number of keystrokes, decrease the user's cognitive load through term recognition (rather than recall) and help the user avoid typing errors (Hendrickson 1989; Jakobsson 1986). It can improve the quality of initial query for known-item as well as exploratory tasks (White and Marchionini 2006; White and Marchionini 2007). RTQE is better than post-query expansion because it lowers task completion time, increases search engagement and increases uptake of interactive query expansion (White and Marchionini 2006). Most research efforts are directed towards improving query expansion suggestions, e.g.(Agichtein et al. 2006; Efthimiadis 2000; Radlinski 2005; White and Morris 2007), and tend to pay less attention to the interface issues. Many RTQE use only *list* organization as presentation style. Prior work has led us to believe that different types of implementation of RTQE presentation would likely result in different user search performance. In (Beaulieu 1997), three different interfaces to the same retrieval system were compared. The study suggests that the quality and effectiveness of search depend on the combination of the retrieval system and its interface in supporting query expansion. Joho et al. (Joho et al. 2002) studied different query expansion presentation styles. They compared two types of organization strategies: alphabetically ordered *list* and *menu hierarchy* IQE interfaces. They found that even though there is no significant difference in the precision-recall between using the two interfaces, people finished the search task significantly faster when using the *menu hierarchy*. Another study (Joho et al. 2004) compared two different hierarchical IQE systems (based on the subsumtion approach and trigger phrased on parent-child description) against a baseline (no suggestion). They found that accessing the hierarchies reduces search iterations, reduces paging actions and increases the chance to find relevant items. In practice, many variants of RTQE organization strategies have been deployed, such as:

- **List** organization strategy, such as by *alphabetical* list (WikiSearch), by past user query in (Kelly et al. 2010), by popular query/destination in (White et al. 2007).
- **Group** organization strategy, such as Google Suggest uses 2 groups: personal

history query and popular prefix match, or `Naver.com` uses 2 groups: popular prefix and suffix match.
• **Menu** organization strategy, such as a cascaded menu hierarchy in (Joho et al. 2002; Joho et al. 2004).
In this research, we focus on the presentation aspects of an autocompletion, namely organization strategies and how they influence users search performance. We are motivated by the usage of relationships of terms from a thesaurus to improve RTQE presentation. Certain relationships between terms from a thesaurus has been known to improve the quality of query expansion. Efthimiadis et al. (Efthimiadis 2000) investigated the terms used in a IQE for the INSPEC database. They reported that variants (synonym) and alternative terms (i.e. narrower, boarder and related terms) relationships are useful for query expansion. Similarly, Joho et al. (Joho et al. 2002) in their research found that for WordNet, the most useful relationships are hyponym, hypernym and synonym. Additionally, they also mention conceptual relation (e.g. teeth-dentist) is a meaningful relationship for query expansion. In this study, we investigate how to improve selection of terms in autocompletion interfaces. In particular, we explore the potential of hierarchical relations in thesauri to improve the organization of autocompletion suggestions. By imposing grouping and ordering strategies we provide a means of navigating the suggestions and finding the right terms faster and easier. We restrict our discussion to interfaces that syntactically complete the input based on exact or partial literal match. We do not consider *query recommendation* which tries to extrapolate queries based on certain (semantic) relations or algorithms, such as in (Bast et al. 2007). Semantic relations are used, if at all, primarily to organize the suggestions in the interface. We carry out two related studies. The first examines the quality of grouping strategies for different thesauri, the second investigates to what extent grouping and (alphabetical) ordering are able to influence the suggestion selection process.

4.2 Organization of Suggestions

In this section, we discuss the look and feel of different organization strategies for autocompletion suggestions used in Study 1 and Study 2 (see Fig.4.1 and Fig. 4.2) Examples are taken for TGN [1] autocompletions, similar visualization and algorithm is applied to WordNet.
Alphabetical order — Fig.4.1a shows autocompletion suggestion in an alphabetical order. The suggestions are organized in the following priority: prefix match

[1] Thesaurus for Geographical Names `http://www.getty.edu/research/conducting_research/vocabularies/tgn/`

on primary literal (location name), prefix match on secondary literal (country name), e.g. the suggestion "*Paris*, Canada" is shown before "*Paris*, France". Exact matches are presented first, followed by partial matches. The first part of suggestion consists of 15 items. When the user selects the "see more" button, all suggestion is presented as a long list.

Group — An organization that combines similar suggestions under a common heading. The grouping category is conveyed visually as a group title. Where terms are related by explicit thesaurus relations, any of these relations can be used as a basis for grouping. Grouping can be based on variants of hyponym relations. There are 2 types of grouping: predefined and dynamic. In predefined grouping the category is always of the same type. For example, TGN's hierarchy is based on geographical containment (e.g. *Europe* > *France* > *Paris*). Grouping can be based on any predefined level within this hierarchy, e.g. grouping by country (Fig. 4.1b). Alternatively, a predefined category can be based on a common property, such as place type (Fig. 4.1c) e.g. inhabited place (city, village) or body of water (stream, lake).

Another variant is the dynamic grouping where the group heading differs and is determined by an algorithm that optimizes groups based on the number of suggestions retrieved. The desired groups can be preset taken from the top level hierarchy (Dynamic TB) or taken from the lowest (leaf) level hierarchy (Dynamic BU). Dynamic groups could provide an alternative grouping for thesauri with irregular hierarchical structures such as WordNet. Fig. 4.1d is an example of Dynamic TB group implemented for TGN.

Figure 4.1: Autocompletion with different organization strategies used in Study 1 for TGN, from left to right: a) alphabetical order, b) Country grouping, c) Place type grouping, d) Dynamic TB grouping

Composite — A composite organization resembles a two level cascaded menu hierarchy. It is groups similar suggestions into a single item (primary menu), deferring their display to a secondary menu. Fig. 4.2d shows an example composite suggestion interface, the primary menu contains all exact match of all location names from the same country. The secondary menu displays more information of the location names that allows disambiguation e.g. *Kingston (the city)* or *Kingston (the parish)*. This strategy retains the simplicity of alphabetical order, but shows larger numbers of alternatives in the limited amount of screen real estate available.

4.3 User Studies

We conducted two user studies to investigate the benefits and trade-offs of different strategies for organizing suggestions in autocompletion interfaces. The first study was an exploratory study to investigate the effects of grouping strategies on two different thesauri. The second study built on the result that grouping terms in a thesaurus of locations can be beneficial. The study investigated different organization strategies on the same set of suggestions.
Technology — The study was done using our autocompletion widget[2] that works on all major browsers supporting (X)HTML, CSS and Javascript. The client side widget is an extension of the Yahoo! User Interface autocompletion widget (YUI v. 2.3.1[3]). The suggestion server has been implemented using SWI-Prolog's web infrastructure[4]. The autocompletion architecture is fully described in (Hildebrand et al. 2007).

4.3.1 Study 1: Grouping Strategies

In Study 1, we investigate different variations of grouping as a type of suggestion organization. As mentioned in the previous section, there are many alternative implementations of grouping strategy using various term relationships in a thesaurus. The goal of Study 1 is to investigate to what extent grouping strategies for autocompletion suggestions can be applied to thesauri and if so, which grouping strategies are meaningful for users. We have chosen to implement similar grouping strategies for two different thesauri: a domain-specific thesaurus, TGN and a global thesaurus, WordNet. Our intention was not to compare the two thesauri, but to evaluate the suitability of different group strategy when implemented for these thesauri. Our research questions in Study 1 are: Can group organization strategy be implemented for the thesauri? Which group strategy is perceived the best by users?

[2] Demo is available at `http://slashfacet.semanticweb.org/autocomplete/demos/`
[3] `http://developer.yahoo.com/yui/autocomplete/`
[4] `http://www.swi-prolog.org/`

Interfaces — We selected 4 autocompletion interfaces to compare with each other for TGN and similarly 4 for WordNet. The four chosen interfaces for each thesaurus are those which we thought were best to offer to users after informal trials of different algorithms and combinations. For TGN, the grouping strategies are: grouping by country (Fig. 4.1b), grouping by place type (Fig. 4.1c) and Dynamic TB grouping (Fig. 4.1d). As a baseline, we used Alphabetical ordering (Fig. 4.1a). We refer these interfaces as the Location Name (LN) interfaces.
For WordNet, the 3 grouping strategies are: predefined grouping using the top nine WordNet category nouns from the hypernym hierarchy, and two dynamic groupings: Dynamic TB and Dynamic BU. Similarly, the Alphabetical order was chosen as a baseline. We refer these interfaces as the Object name (ON) interfaces.
Participants — Participants were recruited by sending out invitations to universities and research institutes from diverse departments, such as computer science, engineering and natural science. In total, 47 people responded. Participants were mostly students and some university employees. All participants reported that they use the Internet daily and are familiar with autocompletion interfaces (e.g. in email clients, search engines and web browsers), 14 participants have experience with autocompletion interfaces in specialized applications such as script editors and interactive script interpreters.
Procedure and Tasks — The study was done as an online interactive experiment. All session activities are logged. Prior to the task, participants answered a short questionnaire focusing about their experiences with autocompletion. Afterwards, every participant was assign tasks with 4 four TGN-LN interfaces (within subject design). For every LN, participants were given the same tasks: to formulate several location queries, such as *Berlin* (city name) or *Alps* (mountain system)), and find the correct location names from the suggestions presented in the interface. Afterwards they were encouraged to try out their own example queries and explore the interface responses. After completing the tasks, participants were asked to answered assessment questions about the quality of the groupings and to give their comments. Finally, participants were asked to rank their preferred strategy for LN, from the most to the least preferred, and provide reasons for their decisions. The same task and procedure were repeated by the participants for the WordNet-ON interfaces. Participants were asked to formulate object queries, such as *Barbecue* or *Party*, and answer the assessment questions about the quality of grouping in this interface. The assessment questions on the quality of the group organization were derived from criteria taken from the literature (Gonzales 1994; Hendrickson 1989; Hodgson and Ruth 1985; Lee and Raymond 1993; Rosch 1988). The answers to the assessment questions are given in a 7-Likert scale (1:low, 7:high). These criteria are:
Q1 - perceived similarity of items within the same group; *"I think the items be-*

longing to each group in this type of list are similar to each other."
Q2 - perceived difference of items between groups; *"I think the items belonging to different groups in this type of list are different from each other."*
Q3 - affinity item and group title; *"I think the relationship between the items and group title is clear in this type of list."*
Q4 - reasonable number of groups; *"I think the number of groups in this type of list is appropriate."*
Q5 - group title appropriateness; *"I think the titles of the groups in this type of list are clear."*
The order of the interfaces were counter balanced using the Latin Square scheme among the participants. Pilot sessions were conducted to ensure that the participants could perform the tasks and understood the questions. The time to complete the study was approximately 30 minutes.
Results — The data we collected from the experiment were processed qualitatively and quantitatively. Our server log indicates that in addition to trying all provided examples, additionally some participants explored the behavior of autocompletion interfaces by trying out their own examples, such as different cities, countries or river names (e.g. Rhein) for LN and various object names (e.g. muscle, mobile, partner) for ON. It is important for us to confirm that participants explore the behavior of the autocompletion beyond the given task before assessing the quality of the interfaces.

Table 4.1: *Top:* Assessment scores (n=47 people, Study 1)

TGN-LN	Mean Score (*SD*) *			
Question	Place type	Country	Dynamic TD	*p-value*
Q1	**5.30(1.68)**	4.57(1.83)	4.34(1.75)	.03
Q2	5.00(1.52)	4.53(1.80)	4.51(1.52)	.71
Q3	5.77(1.49)	5.74(1.51)	5.49(1.57)	.39
Q4	4.91(1.77)	**4.15(1.98)**	4.98(1.76)	.02
Q5	5.30(1.79)	**5.94(1.41)**	5.19(1.85)	.03
WordNet-ON	Mean Score (*SD*) *			
Question	Predefined	Dynamic TD	Dynamic BU	*p-value*
Q1	4.19(1.56)	4.21(1.85)	3.94(1.65)	.77
Q2	4.64(1.47)	4.43(1.60)	**3.96(1.43)**	.01
Q3	4.13(1.81)	4.28(1.75)	4.13(1.66)	.61
Q4	4.19(1.72)	**3.47(1.73)**	4.02(1.88)	.01
Q5	3.83(1.81)	4.04(1.71)	3.72(1.82)	.48

* 7-Likert scale, score 1:strongly disagree, 7:strongly agree

Table 4.2: *Bottom:* Preferred grouping strategy (n=47 people, Study 1).

TGN (LN)	Ease of use (*SD*)	*p-value*
Place type	2.23(1.15)	.16
Dynamic TD	2.35(1.09)	
Country	2.67(1.13)	
Alphabetic	2.74(1.09)	
WordNet (ON)	Ease of use	*p-value*
Alphabetic	1.98(1.23)	.02
Dynamic TD	2.62(.97)	
Predefined	2.68(1.09)	
Dynamic BU	2.72(1.06)	

• *Assessment:* The participants' assessments for six grouping strategies are shown in Table 4.1. We examine each question to understand the characteristics of each grouping strategy using Friedman two-way analysis by ranks[5]. For LN we found that: (a) Place type grouping scored best with respect to perceived similarity - Q1 ($\chi^2(2)$=7.36, p=.03)[6] (b) Country grouping scored best with respect to group title appropriateness - Q5 ($\chi^2(2)$=6.77, p=.03)[7] (c) Country grouping scored lowest with respect to the number of groups - Q4 ($\chi^2(2)$=8.11, p=.02) [8] Perceived similarity indicates the cohesiveness between the suggestions in a group. Place type grouping scores highest for this aspect. Alternatively, the Country group strategy gives most representative group titles (Q5) but scores poor on the number of groups (Q4). One disadvantage of our implementation for the Country group strategy is that we did not make any limitation on the number of groups allowed. Because of this, the autocompletion list can potentially be very long. This is an adjustable parameter of the interface and does not depend on the characteristic of the thesaurus. The assessment score indicates that from the 3 types of LN grouping, Country and Place type are relatively good grouping strategies that each excel in different qualities.
For the ON interfaces, we found that: (a) Dynamic BU group scored lowest with respect to perceived difference - Q2 ($\chi^2(2)$=10.17,p=.01)[9] (b) Dynamic TB group

[5] Nonparametric statistics is used as the data did not meet parametric assumptions

[6] Wilcoxon signed ranks (WSR) *post-hoc* test result for Q1: Place type scored sig. higher than Dynamic TB ($p \ll .05$).

[7] WSR *post-hoc* test result for Q5: Country scored sig. higher than Dynamic TB ($p \ll .05$) and Place type (p=.03)

[8] WSR *post-hoc* test result for Q4: Country scored sig. lower than Dynamic TB (p=.02) and Place type (p=.01)

[9] WSR *post-hoc* test result for Q2: Dynamic BU scored sig. lower than Predefined (p=.01)

scored lowest with respect to the number of groups - Q4 ($\chi^2(2)$=9.66, p=.01)[10] The results showed that none of the ON group strategies excels from each other in the assessment score. We only found that the Dynamic TB and Dynamic BU groups perform the worst in Q2 and Q4. We think this is because the dynamic group strategies actually add to participant's cognitive load when they are trying to go through the suggestion list. No grouping strategy in ON is assessed the best by our participants. The reason for this will be clear in the next results where we compared all group strategies against a baseline (Alphabetical order) and examine users preferences.

• *Preference:* Table 4.2 shows the Mean Rank of each grouping strategy for LN and ON. A low Mean Rank score indicates most preferred, and a high score is least preferred. Using the Friedman two-way analysis by ranks, we found that there is no strong preference in any of the location grouping strategies ($\chi^2(3)$=5.14, p>.05). From the comments provided by the participants, we see that participants prefer different interfaces for different reasons. We conducted the same analysis for the four ON interfaces and found a different result. Participants strongly preferred the Alphabetical order to all other organization strategies ($\chi^2(3)$=10.38, p=.02)[11] From the participants' comments, we understood that they found it difficult to understand the ON grouping strategies. This could explain the strong preference for Alphabetical order.

• *Comments:* Participants' comments gave us an explanation as to their assessment decisions and preferences. It seems that most decisions on choosing a LN interface is based on personal preference. *"By country seems more logical and pragmatic. Place type takes some getting used to but could work fine. Dynamic (grouping) gets confusing, Alphabetical (list is) not very clarifying."* [P6]. For the ON interfaces, participants opinion are more uniform. The main comment was that many participants struggle with understanding ON grouping strategies. The baseline (Alphabetical order) seems to be the easiest to understand based on their past experience with finding terms in a dictionary. *"... I am more familiar with encyclopedic or dictionary structuring. The problem with such group autocompletion advice is that the adaptation process is quite time costly."* [P25].

Retrospective — The main goal of study 1 was to get a feel for how users perceive different grouping strategies. More precisely, we want to find out if and how the different structures of the thesauri used effect the user's perception, and whether the resulting groupings make sense at all.

Ideally, the best grouping strategies are the ones that scores highest on all five

[10] WSR *post-hoc* test result for Q4: Dynamic TB scored sig. lower than Predefined ($p \ll .05$) and Dynamic BU (p=.03)

[11] WSR *post-hoc* test result for Mean Rank of preference: Alphabetical scored sig. lowest (i.e. strongly preferred) then Predefined (p=.02), Dynamic TB (p=.04) and Dynamic BU (p=.01).

assessment scores (Table 4.1) and most preferred (Table 4.2). However, this is not the case. For LN, we found that one grouping strategy is better in some aspect while others in another aspect. For ON, we did not found any favoured grouping strategy.
Thus, we concluded that grouping strategies may not be suitable for every type of thesaurus. For a domain-specific thesaurus, such as TGN we could find a sensible grouping strategy that people could understand relatively easily. In a global thesaurus such as WordNet, however, we found different results. The users preference, assessment scores and participants' comments lead to the conclusion that for WordNet the group organization may not be the best strategy to use. In cases where the underlying thesaurus does not provide the information necessary for appropriate grouping, Alphabetical ordering is the best option.

4.3.2 Study 2: Organization Strategies

Based on what we have learned in Study 1, we conducted a follow up study. We narrowed down the scope of study 2 by only investigating autocompletion for TGN. We decided not use WordNet since none of our group strategies for WordNet outperformed the baseline (Alphabetical). The goal of the second study is to compare three types of autocompletion: Alphabetical order, Group and Composite. We would like to investigate which interface helps users to search for terms from a thesaurus the fastest and easiest. To be able to come to this conclusion, we setup an experiment where users are required to use autocompletion for known-item search tasks. To evaluate speed, we measure performance in time to complete task (objective measurement). To measure ease-of-use, we took three subjective measurements: user assessments, preferences and comments. Additionally, we analyze the quality of keywords provided in each condition.
Interfaces — In this study, we compared 4 different interfaces, namely: Alphabetical order (Fig.4.2b), Group (Fig.4.2c), Composite (Fig.4.2d) and a no autocompletion (NAC) interface (Fig.4.2a).
Participants — We recruited participants in the same manner as in the first study. In total, 41 people participated. Participants were aged between 16-66 years (M=30.90, SD=10.45). In general, participants use the Internet frequently (M=34.96, SD=19.51) (hours per week), and have medium to high familiarity with autocompletion interfaces[12].
Procedure — Each participant is assigned interfaces: NAC, Alphabetic, Group and Composite (within subject design). The order of the conditions were counter

[12] 1:low familiarity, 5:high familiarity; Autocompletion in search engines (M=3.40,SD=1.34), email client (M=4.12,SD=1.17), address browser(M=3.86,SD=1.46), Misc.: autocompletion in MS Visual Studio, Eclipse IDE

balanced using the Latin Square scheme among the participants. Pilot sessions were conducted to ensure that the participants could perform the tasks and understood the questions. The time to complete the study is approximately 25-30 minutes. In the experiment, participants started by answering general questions about their experience in using the Internet and autocompletion. Participants were then given a trial session to get accustomed to the interfaces. During the experiment, participants were given 24 tasks. In every task, time measurements were taken and participants were asked to assess the usability of the interface afterwards. We are interested in comparing the usability aspects of the different interfaces. After every interface, participants answered two questions (5-Likert scale):
Q1 - *"I find this interface easy to use."*
Q2 - *"I find the organization of the suggestions easy to understand.*"

At the end, participants were asked to rank the autocompletion interfaces based on their preference and to give reasons for their choices.

Task — Participants were given 24 tasks (3 tasks per interface). To simulate a realistic search task, participants were asked to search and specify the birth place of a famous person (see Fig. 4.3). They were allowed to find the answers in Wikipedia and then fill in their answer using the autocompletion interface. Participants were encouraged to use autocompletion but could choose not to use it if they could not find the right suggestion from the list. We have chosen the non trivial tasks such as locations with exactly the same name. Thus, for all questions, the need for for disambiguation and choosing the correct terms was clear. For example, the birth place of Kurt Kobain (Aberdeen, Washington, USA) has at least 56 other similar place name matches, of which Aberdeen in the UK will most likely be the most familiar to our European participants. The times recorded are the autocompletion typing time only. We disregard the time it takes for the participant to browse the Web and look for answers.

Results

• *Mean keystrokes:* As expected, an autocompletion interface reduces the number of keystrokes required to type. On average, users typed almost twice as many characters in the NAC condition compared to when using autocompletion (see Table 4.4). Additionally, we found that some participants copied and pasted the location name they found from Wikipedia. This behavior was identified from the interaction event log and is estimated to be about 7.5% from the total tasks performed by all participants.

• *Performance in Time:* Table 4.4 shows the mean time it took for participants to complete a task. This time constituted the time from the first keystroke typed until

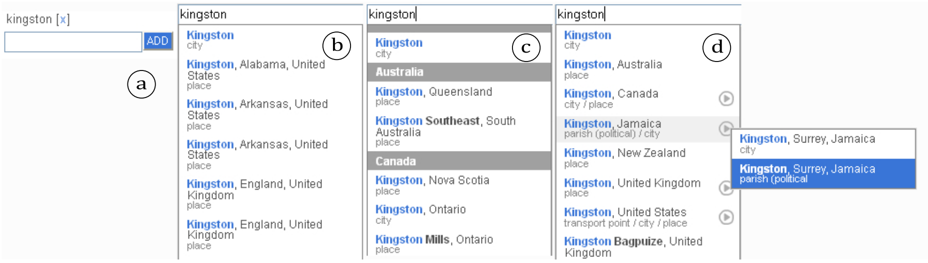


Figure 4.2: Autocompletion in Study 2, from left to right: a) NAC, b) Alphabetical order, c) Group by country and d) Composite.

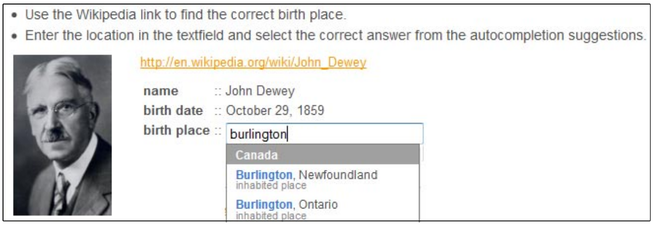


Figure 4.3: Task example used in Study 2

selecting a suggestion (for the autocompletion conditions) or hitting the return key (for the NAC condition). When we compare the performance of the individual autocompletion interfaces, we find that Group and Composite are significantly faster (47% and 45% resp.) than the Alphabetical order[13] We conclude that both Group and Composite strategies help the user search for terms faster than Alphabetical order.

• *Quality of keywords:* Table 4.3 shows the quality of keywords provided by participants. The quality of keywords is measured by how accurately the location names are given. We found four levels of accuracy (from low to high): strings that consist of one term (mostly city names which are ambiguous because there exist many cities the same name, e.g. *"Kingston"*), strings that consist of two terms (mostly a combination of city and state, or city and country, e.g. *"Kingston, USA"*), strings that consist of three terms (mostly a combination of city, state and country names, e.g. *"Kingston, Texas, USA"*) and keywords which were correctly chosen from the

[13] WSR *post-hoc* test result for Time: Group is sig. faster than Alphabetical ($p \ll .05$). Additionally, Composite is sig. faster than Alphabetical ($p \ll .05$).

Table 4.3: Quality of keywords provided by participants (492 tasks, 41 people, Study 2)

Interface	NAC	Alphabetical	Group	Composite
Total correct keyword	96.7%	86.2%	95.1%	84.5%
a. Unique concept	n/a	**77.2%**	**86.2%**	**82.9%**
b. One term	14.6%	2.4%	0.8%	0%
c. Two terms	**53.7%**	6.5%	5.7%	0%
d. Three terms	28.4%	0%	2.4%	1.6%
Total incorrect keyword	3.2%	13.8%	4.9%	15.4%
a. Select wrong item	n/a	13.0%	4.9%	15.4%
b. Typing error	**2.4%**	0%	0%	0%
c. No answer	0.8%	0.8%	0%	0%

Table 4.4: User search performance and preference (492 tasks, n=41 people, Study 2)

Interface	NAC	Alphabetical	Group	Composite
Mean no of keystrokes (*SD*)	**19.20(6.86)**	8.55(4.50)	7.89(4.81)	7.91(3.82)
Mean time in s (*SD*)	5.94(3.41)	38.93 (46.87)	18.36 (10.99)	17.62 (12.25)
Mean pref. rank (*SD*)	2.93(1.23)	2.71(.90)	**1.98(1.11)**	**2.39(1.02)**
Mean score Q1 * (*SD*)	3.07(1.21)	2.59(.87)	**3.34(1.39)**	**3.56(.90)**
Mean score Q2 * (*SD*)	n/a	3.05(1.24)	**3.73(1.10)**	**3.61(.95)**

* 5-Likert scale, score 1:strongly disagree, 5:strongly agree

suggestions (unique concepts from the thesaurus). All autocompletion interfaces have a high percentage of correct keywords (all above 84.5%). The quality of keywords provided by participants, however, differs when using autocompletion and without. In NAC, most keywords consist of merely 2 terms (53.7%), which is in many cases not sufficient for location name disambiguation. For example, there are 47 places named *Kingston* in the USA. Only about a third of the cases in NAC (28.4%) consist of three terms. In contrast, keywords provided in the other autocompletion interfaces are mostly high quality keywords that are unique concepts (86.2% for Alphabetical, 95.1% for Group, and 84.5% for Composite). We also identified three types of incorrect keywords: incorrect terms selected from the autocompletion suggestions, typing errors and blank entries where participants gave no keyword at all. A closer look at the incorrect keywords provided by participants reveals that most errors in the NAC conditions are typing mistakes (2.4%), while most errors in the autocompletion interfaces are wrong autocompletion selection. For example, selecting *Ottawa (the river)* instead of *Ottawa (the city)*. From all 3 autocompletion interfaces, Group organization generates least error (4.9%) compared to Composite (15.4%) and Alphabetical (13.8%). The results show that even though NAC is slightly faster, the quality of keywords provided in any of the autocompletion interfaces are far higher.

• *Perceived ease-of-use:* We gathered participants assessments on the ease-of-use of each interface (see Table 4.4). In general, people find the Group and Composite interface easier to use than Alphabetical and NAC interface ($\chi^2(3)$=17.52, $p \ll .05$)[14] In a follow up question (Q2), we wanted to know specifically if people understood the organization strategy. Most people agree with the statement that Group and Composite suggestion organization is easier to understand than Alphabetical list ($\chi^2(2)$=8.12, p=.02)[15]. We conclude that both Group and Composite interfaces are perceived easier to use and understand than the Alphabetical order.

• *Preference:* Our analysis shows there is a preference for Group strategy (see Table 4.4), although Composite is not far behind ($\chi^2(3)$=12.6, $p \ll .05$)[16]. From the comments made by the participants we understand more about the reasons behind the users preference. Participants acknowledge that autocompletion suggestions help avoid typing mistakes and enable them to express more keywords than they would otherwise have thought of. *"The lack of autocompletion choices prevents me to give a proper answer for question X."*[P1]. In general, participants think Group organization is better. *"It's comfortable to see the countries separated"* [P2]. *" You know where you have to go. You get a better overview"* [P16]. For many participants, the Composite organization is relatively new. The main disadvantages of Composite are: a) requires more interaction with the interface before making a selection (e.g. mouse movement and click) and b) submenu interaction requires getting used to *"took several seconds to discover the small arrows. After that, the interface is easy to use"* [P4]

4.4 Discussion

Study limitation — In Study 1, the grouping strategies tested were developed partly by trial and error in combination with educated guesses. We came up with a range of possible grouping strategies, which we tested informally. We only formally tested the three different grouping strategies which performed best in the informal test. The result of Study 1 shows that, in contrary to the TGN grouping strategies, our best grouping strategies for WordNet were not helpful for users and people pre-

[14]WSR *post-hoc* test result for Q1: Group is sig. perceived easier-to-use than Alphabetical ($p \ll .05$). Composite is sig. perceived easier-to-use then Alphabetical ($p \ll .05$). No difference between Group and Composite.

[15]WSR *post-hoc* test result for Q2: Group organization is sig. perceived easier to understand than Alphabetical (p=.01). Composite organization is sig. perceived easier to understand than Alphabetical (p=.04). No difference between Group and Composite.

[16]WSR *post-hoc* test result for preferred interface: Group organization is sig. preferred than Alphabetical ($p \ll .05$) and NAC ($p \ll .05$). No sig. difference between Group and Composite.

ferred Alphabetical ordering as an organization strategy. We acknowledge that it might be the case that we did not succeed to find the appropriate grouping for WordNet. Therefore it is reasonable to only conclude that for WordNet grouping strategies which were tested, none outperformed the Alphabetical list.
During the experiment, we observed that there are some cut and paste behavior specially for the NAC interface. This might provide additional explanation as to why participants complete NAC tasks faster than other tasks. We expect that if all participants were only allowed to type (not cut and paste), the autocompletion interfaces would show comparable time performance.
Alphabetical order — When using a global thesaurus, such as WordNet, Alphabetical order organization seems to be the best option. This organization requires very little learning effort. The downside of this organization is, as one participant points out, that it provides no "overview" when there are many suggestions.
Grouping strategy — We learned from the first study that a grouping strategy should be chosen carefully because not every grouping strategy is suitable to use. The TGN groupings produced by the thesaurus hierarchy seem to be more natural than WordNet groupings. In many of our pairwise statistical comparisons between Group and Composite organization, we found no significant differences. Based on this study alone, we cannot see a clear advantage one type of organization over the other. We can say, however, that the Group organization has a tendency to expand the length of suggestion interface vertically, whereas the Composite organization has a tendency to reduce the length of the suggestion interface. Therefore, depending on the thesaurus used and the length of suggestions it produces, the Composite organization might have an advantage.
Improve autocompletion — The server log indicates that some people use commas and make keystroke errors. We learned that in order to make a good autocompletion interface, there are a number of supporting functionalities that are indispensable: (a) Compensate for non alphanumeric letters such as white space(s) and commas. For example, the system should know that *Kingston Jamaica* is the same query as *Kingston, Jamaica*. Our finding is consistent with the study in (Teevan et al. 2007) on how people express similar queries in different ways. (b) Users may make typing mistakes (e.g. *Ottawa, Ottowa, Otawa*). Spell check and giving suggestions based on likely spelling would be a useful feature.
Incentive to use autocompletion — We observe that autocompletion can stimulate people to be more precise in their keywords. Even though participants were instructed to be "as precise as possible", the keywords provided in the no-autocompletion tasks are largely ambiguous: only less than a third of the keywords consist of detailed information (i.e. city, state and country). This is in large contrast to the keywords provided by the participants when using autocompletion suggestions where they are mostly unambiguous location concepts. Autocomple-

tion allows people to provide high quality keywords from which an information retrieval system can benefit. We believe that users are willing to spend more time to formulate a more elaborate query with the help of autocompletion interfaces if the option to use is made available.

4.5 Conclusions and Future work

We conducted serial user studies to compare different kinds of organization for autocompletion suggestions that can help improve known-item search task. In the first study, we found that grouping strategies might not be suitable when using a global thesauri, such as WordNet and only certain grouping strategies could be used for TGN. Based on what we have learned, we conducted a second study where we compared three different autocompletion suggestion interfaces. In general, we found that the quality of keywords provided by users are better with the autocompletion; Group and Composite organization help users search faster than when using the Alphabetical order; users perceive Group and Composite easier to use and to understand. We are currently integrating autocompletion with our applications and evaluate its performance for a domain-expert annotation task. In addition to this, we will improve autocompletion interface to detect similar query strings identified in (Teevan et al. 2007), such as synonyms, extra whitespace and word swaps.

CHAPTER 5

Designing a Thesaurus-based Comparison Search Interface for Linked Cultural Heritage Sources

We investigate support for comparison search, a specific type of information gathering task (research question 3). As indicated in earlier chapters, this task occurs frequently for both experts and lay users. We develop a thesaurus-based comparison search tool to support cultural heritage experts. The tool allows a user to search, select and compare sets of artifacts from multiple sources. Different visualizations are available, allowing users to use different comparison strategies to cope with the underlying heterogeneous data and the complexity of the search tasks. We carry out two user studies. In the preliminary study, we identify requirements for comparison search tasks (research question 2). In the second study, we examine the effectiveness of the tool in helping to solve comparison search tasks. Our main contribution here is to establish design guidelines for the data and interface of a comparison search application. Moreover, we offer insights into when thesauri and metadata are appropriate for use in such applications.

This chapter was published as "Designing a thesaurus-based comparison search interface for linked cultural heritage sources" in the International Conference on Intelligent User Interfaces (IUI'10) (Amin et al. 2010) and was co-authored by Michiel Hildebrand, Jacco van Ossenbruggen and Lynda Hardman.

5.1 Introduction

In an in-depth study on information seeking needs in the cultural heritage domain, comparison search was identified as an example of an information seeking task that experts perform frequently for their work (Amin et al. 2008). Comparison search involves examining objects or sets of objects for similarities and differences. The more objects there are to compare, and more properties to compare on, the more complex the task becomes. For such tasks, support tools are indispensable. Comparison search tasks are commonly found in the e-commerce domain. For example, a customer who wants to buy a product online might be interested in comparing products from different manufacturers using various properties, such as quality, price, features, and delivery time.

Comparison search also occurs in the cultural heritage domain. For example, consider an art historian doing comparative studies on Dutch paintings owned by different museum collections throughout the Netherlands. First, the historian needs to thoroughly *search* for all paintings made by Dutch artists in different archives and digital museum collections.

Second, after all artworks are *selected*, the historian might need to *compare* the artworks by their distinguishing properties, such as by artist, materials used, art style or year of creation, to be able to identify trends.

In practice, the scenario above is not trivial. First, there are difficulties in thoroughly searching for artworks in collections. Many museums have their own thesauri from which they use terms to annotate their artworks. As a result, a user needs to be familiar with the different terms in these thesauri and how they are used to annotate the artworks. This is unlikely, as typically, only the museum employees have such detailed knowledge. Second, most museum collections and archives offer poor support for complex search tasks such as comparison search. Most tools used only offer simple interaction, such as keyword search. This is found to be too restrictive for complex tasks such as comparison search (Amin et al. 2008).

In this chapter, we discuss a user-centric approach to support cultural heritage experts in searching and comparing artworks. This research consists of several phases. We carried out a preliminary study to better understand how experts conduct comparison search in practice for their daily work. Based on this study, we derived design requirements and identified key features for a thesaurus-based comparison search tool that supports users to search, select and compare artworks. We then implemented LISA to help experts compare sets of artworks. Finally, we evaluate how well LISA supports the comparison search task. The contribution of this chapter is to establish design requirements for the data and interface of a comparison search application. Moreover, we offer insights into when thesauri and metadata are appropriate to use for such applications.

5.2 Related Work

To the best of our knowledge, no other tools currently support comparison search in the cultural heritage domain. We first discuss the state of the art information access tools in cross-collection cultural heritage search sites followed by support for comparison search in other domains.

As a result of many digitization projects, the collections of many museums, archives and libraries are now accessible online. Recently, new aggregated search applications allow users to search directly in multiple collections. Examples include portal sites such as `www.europeana.eu` and `www.collectiewijzer.nl`; and research prototypes such as MultimediaN E-Culture (Schreiber et al. 2008) and CultureSampo (Hyvonen et al. 2009).

Such systems are useful, not only because they enable users to find pieces of information faster, but because of their potential for comparing objects from different collections. Most of these systems include interfaces for common information seeking tasks, namely searching, browsing and exploration (Shen et al. 2006). Unfortunately, however, the interfaces of none of the systems mentioned above support comparison tasks directly. Rather, experts are forced to use the standard search and browse interfaces provided.

Related work on comparison search can be found, however, in other domains. For e-commerce applications, analysis of requirements for supporting comparison search has been well covered in the literature (Callahan and Koenemann 2000; Kobsa 2001; Lee et al. 2004; Steiger and Stolze 1997). One important requirement of a product comparison search interface is to allow users to extensively search and browse objects before comparing (Callahan and Koenemann 2000; Steiger and Stolze 1997). Selection is also another crucial requirement and can be done in different ways. For example, a study on electronic catalogs (Steiger and Stolze 1997) emphasizes the importance of allowing users to conduct incremental object selection in comparison search. Another example is to use interactive object filtering based on the available properties, such as in VOPC (Lee et al. 2004). In this interface, the properties are visualized next to each other, providing the user with an overview of all possible properties to select from.

A visualization requirement for comparison search is the capability to present multiple properties (Lee et al. 2004; Steiger and Stolze 1997). Different visualisations have different characteristics. Some presentations are able to show values of a single property (e.g. Bar charts, Dotplots, Histograms and Spinograms), two properties (e.g. Scatterplot, Mosaicplot), or multiple properties (e.g. Table) (Theus and Urbanek 2009; Young et al. 2006). Bar charts are simple and straight forward. All values of a single property can be displayed in ascending or descending order. Scatterplots are mostly used when two important characteristics need to be displayed at once, such as in Gapminder `www.gapminder.com` and in (Perugini et al. 2004; Plaisant et al. 2006). Tables are still the most popular visualisations for comparison search in practice. Tables present information in a simple way: the products are presented all in one column, while the product properties are presented in a row (or vice versa) (Spenke et al. 1996; Tenev and Rao 1997). This type of presentation allows a user to clearly see the values for multiple properties for multiple products at the same time.

The effects of different visualisations on user performance have also been studied, e.g. in (Callahan and Koenemann 2000; Kobsa 2001). The experiment in (Kobsa 2001) shows that table-like interfaces help users solve problems faster, while a scatterplots is better at guiding users find correct answers. Callahan et al. (Callahan and Koenemann 2000) show that an interactive table (InfoZoom) helps user compare object properties faster than a hierarchical table. The hierarchical table interface, however, was found to be more pleasant to use. The study also suggests that the user's performance, while using a comparison search interface, depends on the type of task, the context and the ability of the user to translate the given problem while working with the system (Callahan and Koenemann 2000). Thus, different domains and tasks might have different comparison search requirements.

5.3 Preliminary Study: Understanding Comparison Search in the Cultural Heritage Domain

We conducted the preliminary study with two goals in mind. First, to identify problems that experts face when they conduct comparison searches. Second, to derive realistic use cases about comparison search tasks that cultural heritage experts carry out during their work.

5.3.1 Setup

We carried out one to two hours semi-structured interviews that took place at the participant's working environment. Each interview consisted of several parts, starting with an introduction explaining the study and general demographic questions. We then asked questions related to accessing multiple cultural heritage sources and how to compare results coming from these sources. Next, we showed sketches of a hypothetical comparison search prototype. These were displayed on the computer to animate interactions with the interface. Finally, the participants were given the opportunity to ask any questions or address concerns about the sketches. In total, seven cultural heritage experts from three different institutions took part in the interviews (see Table 5.1). The participants' average age was 39 years old. They had diverse roles: 2 researchers, 3 curators, 1 art historian and 1 consultant. Most participants had senior positions and had a good overview of the different expert roles within their organization. We hoped that our participants would thus be able to provide insights into the work of their colleagues as well as into their own. All interviews were voice recorded for documentation.

5.3.2 Results

We divide the participants' comments into two themes: the comparison search task use cases that experts conduct for their work, and challenges that experts face in conducting these tasks.

5.3.2.1 Comparison search use cases

In the interview, we asked participants for instances of comparison search tasks that they, or people in their community, would conduct as part of their daily work. To stimulate ideas, we showed them mock-ups of a comparison search tool. Based on the comparison search demonstrated in the mock-up, participants described several use cases along with the roles of those involved.

Learning about collections — As a part of their education program in art history and museology, students are required to familiarize themselves with the variety of museum collections. Currently, this can only be done by browsing through different museum websites individually. *"Students study sculptures from different museums to get a first impression about what different museums have related to sculptures."* [P6]

Planning exhibitions — Whenever a curator needs to prepare for an exhibition,

Table 5.1: Preliminary Study: Participants demography (total: 7 people)

Age:	35-42 years old (M=39.3, SD=2.8)
Gender:	1 male, 6 female
Affiliation	CH institution(6), museum(1),
Expert role:	researcher(2), curator(3)
	art historian(1), consultant(1)

s/he needs to find and collect different artwork candidates. This is followed by a selection process where the curator compares and judges each artwork to find the most suitable ones to be displayed in the exhibition. *"I'm preparing an exhibition on [a painter], and need to make a selection and then from the selected items decide which ones should be finally picked for the exhibition."* [P4]

Museometry[1] — Museum collections change throughout time. Artworks may be loaned, borrowed, sent for restoration or donated. The museum management needs to have periodic quantitative reports of the distribution of artworks to allocate appropriate amount of resources, e.g. for risk management or for expert training courses. *"A museum with diverse collections will benefit more from this interface [mock-up], in particular from the managerial perspective. If I can see, for example, that 80% of my collections are [made of] wood, then I know how many resources to allocate for wood preservation."* [P2].

Qualitative comparison — Experts often need to conduct qualitative comparisons on other experts' assumptions, opinions or recommendations. This task may require in-depth cognitive analysis and interpretation. For example, one participant often needs to analyze different point of views *"I would like to compare arguments between experts about a particular topic. First look at a lot of projects and look for best practices. Always choose from internationally recognized studies."* [P3]

5.3.2.2 Challenges in conducting comparison search tasks

The challenges mentioned by participants were primarily about two issues: searching and specifying the terms (e.g. name aliases, multiple languages and multiple terms), and comparing multiple sets and multiple properties. Descriptions of these issues are presented below.

Name aliases — The participant may not always specify the correct name when searching because s/he does not know which variant is used in the collection. For example, location names change with time, e.g. Burma (old name), Myanmar (official name). Artists may also have different name aliases. *"The problem is you don't always know how to write the artist name that belongs in a specific collection."* [P2]

Multiple languages — A related problem is when artworks have multilingual annotations. Artworks coming from all parts of the world may be annotated in their vernacular terms or other languages (e.g. Spain, Spanje, España). In order to find these artworks, traditionally, the user needs to perform multiple searches using all possible terms and languages. Not only is this task tedious but also not

[1]Museometry: research that emphasizes a quantitative approach to answer questions related to different aspects of museum information and its quality.

always obvious for users. *"How could I search for artworks if the language is different?"* [P5]
Multiple terms — There are many potential terms that museums can use to annotate their work. To an outsider, even with some level of domain expertise, guessing which search terms to use is not obvious. One museum curator mentioned that she often needs to help website visitors with their searching. *"Sometimes visitors (of the museum collection website) do not know what to type. For example, to search for an Islamic collection, there are many different words: Islam, Islamic, Moslem, Muslim. Thus sometimes I do the searching for them and send the (search result) link."* [P6]
Comparing many sets — The tools used do not support the comparison of multiple sets of artworks e.g. comparing the differences between artworks from different museums. For example, one participant researching on museum management usually compares 77 contemporary art museums in the Netherlands at the same time. *"Most of the time I compare more than one museum."* [P6]. Comparison may be based on various characteristics, such as artworks from different museums or artworks from different artists.
Single and multiple property comparison — Tools do not support comparison search tasks using multiple properties: *"Compare collections of artworks by female artists from before and after the 1960's."* [P1]. In another example, the curator wants to highlight different aspects of African art collections in museums in the Netherlands, and wants to compare how different cultures (e.g. Akan, Gurma) predominately create different artwork types (e.g. mask, painting) and how this changes according to the history of the nation (e.g. pre-colonialism v.s. post-colonialism).

5.3.3 Key findings

The preliminary study provided us insights into comparison search tasks conducted by cultural heritage experts. Our intention was to see where a thesaurus-based comparison search tool can help the experts. Comparison search use cases, such as learning about collections, planning exhibitions and museometry, may require quantitative processing of artwork's metadata. Since there can be many homonyms and synonyms within the metadata, simple text matching is insufficient. Quantitative computation on the artwork's metadata taking the thesaurus that provides the metadata terms could yield more accurate results. Therefore, for these use cases, a thesaurus-based comparison search tool maybe useful. For other use cases, such as qualitative comparisons, a thesauri-based comparison tool might not be sufficient. Qualitative comparison requires in-depth analysis and interpretation of information primarily coming from non structured data, such as literature and articles. Thus, for these use cases, metadata and thesauri are insufficient information sources. This is confirmed by participants. For example, when asked about using metadata for comparing different experts' points of view, participants commented: *"Sometimes the information is just too basic."* [P5], *"This is not specific enough for the job."* [P2]

Furthermore, we found that there are three main challenges in comparison search: (1) to support searching using the terms that match with the object's metadata; (2) to select the objects to compare and (3) to support comparison for multiple sets and properties.

5.4 Design Requirements

The preliminary study and related work gave us directions on the design requirements for a thesaurus-based comparison search tool for the cultural heritage domain: it should provide features that help users overcome the search, selection and comparison challenges:

Searching artworks — Most of the problems when searching for artworks, such as name aliases, multiple languages and multiple terms, are related to finding terms that match with the artwork's metadata. To help users find the matching terms, a *guided search*, such as an interactive query expansion interface (Amin et al. 2009), feature should be provided.

Selecting artworks — Selecting artworks is an important part of a comparison search task (Lee et al. 2004; Steiger and Stolze 1997). Selection is an intermediate step where participants define a set of artworks to be compared against each other. While this activity is enabled in some museum collection websites that cultural heritage experts frequently use (RKDimages, `http://english.rkd.nl/Databases/RKDimages`), the interface and interaction is often unintuitive and sluggish. There are two important requirements for the artwork selection process: first, the selection process has to be easy and convenient; second, to ensure a smooth selection process, the user has to be able to add and to remove any artwork from a set.

Comparing artworks — There are two requirements for this activity: first, a comparison search tool for the cultural heritage domain should support comparing multiple sets, where a set can contain many artworks. In the presentation, the differences and similarities between these sets should be clearly distinguishable. Second, it is important to support a comparison on single and multiple properties. To focus our research, we concentrate on supporting one property (single property) and two properties (dual property) comparison. For different type of comparison search tasks (single and dual property) a suitable presentation type should be made available (Theus and Urbanek 2009; Young et al. 2006).

5.5 LISA: Design and Implementation

The LISA application[2] is part of a suite of tools developed within the MultimediaN E-Culture project[3]. The project concentrates on providing intelligent access to distributed and heterogeneous cultural heritage collections. In the following sections we discuss LISA's technology infrastructure, the datasets and the user interface. We focus our discussion primarily on the interface and interaction. For an extensive discussion on the technological infrastructure, see (Schreiber et al. 2008; Wielemaker et al. 2008).

5.5.1 Design

Overview — To support the identified activities of comparison search, the LISA interface consists of four areas (see Fig.5.1): (a) the search area, (g) the search result area, (k)(l) the selection areas, and (j) the comparison area. In the search area, the user incrementally formulates queries, and the results are shown in the search result area below. In the selection area, users can specify and see two sets of artworks. In the comparison area, there are alternative visualizations that can be used to see the characteristics of the selected artworks. We discuss the three supported activities separately. As an example, we compare the self-portraits of Vincent van Gogh with the self-portraits of Rembrandt van Rijn (see Fig.5.1).
Search — To support search, we use a thesaurus-based guided search that consists of a property filter and autocompletion in the interface. The property filter is a pull down menu that shows all possible artwork properties, such as artwork creator, creation date or material, that can be selected (see Fig.5.1b). When a user types a keyword, the autocompletion interface will show suggestions of terms used by the museum collections. An alphabetical ordering is used for our autocompletion suggestions as we found it to be the most effective ordering for loosely structured thesauri (Amin et al. 2009). The guided search takes the form of `property:value` pairs. Users can complete a search by selecting an autocompletion suggestion (Fig.5.1d). To assure full flexibility, the user can add and remove as many `property:value` pairs as s/he wishes. In the example of Fig. 5.1, to search for self-portraits of Rembrandt van Rijn, the user specified two guided searches: to search for all artworks having a

[2]The LISA prototype in progress is accessible at: `http://e-culture.multimedian.nl/lisa/compsearch`.
[3]`http://e-culture.multimedian.nl/`

`Creator:Rembrandt van Rijn` *and* `subject:zelfportret` (English: self-portrait) (see Fig 5.1c), and to search for all artworks having a `Creator:Vincent van Gogh` *and* `subject:zelfportret`.

Because related terms from different thesauri are linked, the system can provide a match even though the artworks' metadata are different as long as they are semantically equivalent. For example, there are 31 name aliases for 'Rembrandt van Rijn' in ULAN[4], e.g. 'Rembrandt van Ryn' or 'Rembrandt van Rhijn'. If a user specify any of these alternative names, LISA will be able to retrieve the same artworks. Similarly, it is possible to thoroughly search with different geographical, art and architecture, and iconographic name aliases. For a list of all linked thesauri supported by LISA, refer to Table **??**. For more information about designing and configuring a thesaurus-based autocompletion see (Amin et al. 2009) [5].

Selection — After the user is satisfied with the search results, s/he needs to select the artworks to compare. The system allows the user to add multiple artworks in either set. In the current implementation, we only support comparison between two sets [6]. An artwork can be placed in any of the available sets (Set A or Set B). There are two easy ways to add an artwork to a selection: first, by dragging and dropping an artwork thumbnail from the search result panel to the selection panel (Fig. 5.1k,l); second, a bulk selection of all search results can be made by clicking the *Set A* or *Set B* button (Fig. 5.1f). To allow fine tuning of selections, adding or removing an artwork from the set is made possible. The search and selection process are typically done sequentially. For example, first the user searches for all self-portraits of Van Gogh (`creator: Vincent van Gogh`, `subject: zelfportrait`), places them on the Set A selection area (Fig. 5.1k). Afterwards s/he makes a second search of all self-portraits of Rembrandt (`creator: Rembrandt van Rijn`, `subject: zelfportrait`) and places the results in the Set B selection area (see Fig. 5.1l).

Comparison with visualizations — LISA currently supports single property comparison and dual property comparison. We choose the Bar chart (Fig. 5.1j3) for single property comparison and the Scatterplot (Fig. 5.1j1) for dual property comparison because these presentations are the most common from a variety of visualizations specified in (Theus and Urbanek 2009; Young et al. 2006). Additionally, we also implemented a Table visualization because this type of presentation is

[4]Union List of Artist Names thesaurus, `http://www.getty.edu/research/conducting_research/vocabularies/ulan/`

[5]In this implementation, only syntactic matches are shown in the suggestions. This is, however, configurable to also suggest semantic matches, such as broader, narrower and related terms

[6]This is extendable as the current application design takes into account the future addition of more sets, however, since computations are carried out client-side, for more than 1,000 objects the interface becomes slow.

the default presentation for most comparison search applications (see Fig. 5.1j2). Whenever an object is placed in the selection area (Set A or Set B), the visualization area is updated. In all visualizations, we use color codes to indicate which sets the elements belong to (either Set A or Set B). Fig. 5.1j3 shows a Bar chart representation of artworks from both sets. The bar chart highlights the comparison between the two sets with respect to the chosen property:`material`. The x-axis represents the artworks organized by the selected property:`material` in alphabetical order. The y-axis represents the number of artworks. The figure shows the values of Set A and Set B next to each other. The property pull-down menu (Fig. 5.1o) shows all available properties for which the objects can be organized, e.g. by dimension height, date, material or depicted subject. The scatterplot presentation (Fig. 5.1j1) shows comparison of sets with respect to the dual property selection (Fig. 5.1i) i.e. `material` (y-axis) and `date` (x-axis). This presentation highlights the differences between the two sets with respect to the creation time and what materials they are made of. The table visualization (Fig. 5.1j2) shows all artworks from Set A. To view all properties from Set B, the user needs to select the tab (Fig. 5.1m). It is possible to explore the information space by two means: a) alternating between different properties by selecting an item at the property pull down menu (Fig. 5.1o), or b) visual exploration either by zooming, panning or scrolling. With the different visualizations, the user can flip through different properties to examine multiple collections simultaneously to gather quick insights about vast collections, which is extremely difficult with current tools.

5.5.2 Implementation

Infrastructure — The LISA application is developed on top of *ClioPatria*, a web application platform for search and annotation across heterogeneous collections. For detailed information on the web server infrastructure and the search strategies across heterogeneous collections, see (Schreiber et al. 2008; Wielemaker et al. 2008). Communication between the client and the server is done via requests to the system's HTTP API. Information is sent back from the server in JSON. The implementation of the interface uses (X)HTML, CSS, Javascript and Flash. It is tested on the Firefox 3.0.10 browser. The client side visualization widgets use an extension of the Yahoo User Interface widget (YUI v. 2.7.0) and amChart v. 1.6.5 [7].

Dataset — To enable comparison search with LISA, the server needs to host common thesauri, namely IconClass[8], the Getty Art & Architecture Thesaurus

[7] `http://www.amChart.com`

[8] `http://www.iconclass.nl`

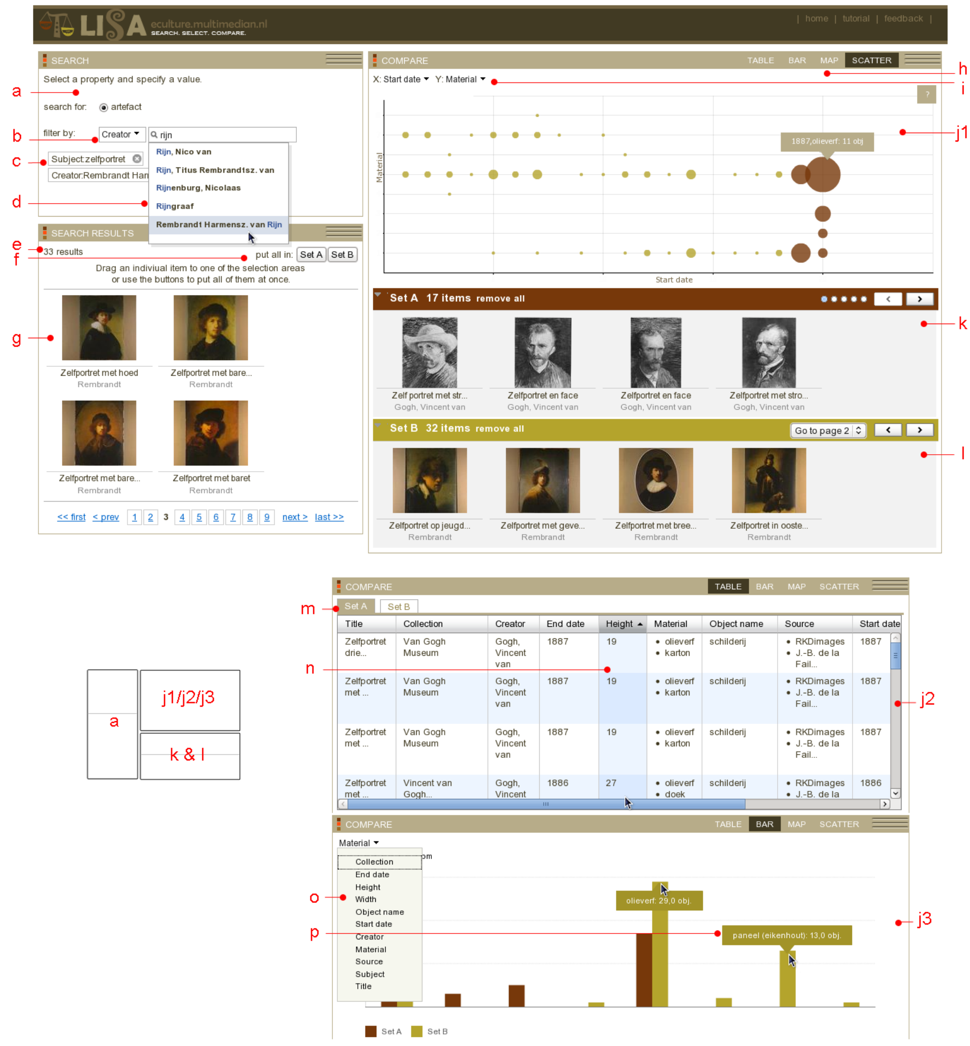


Figure 5.1: The LISA interface (a) search area, (k) selection area Set A, (l) selection area Set B, (j) comparison area: (j1) Scatterplot (j2) Table (j3) Bar chart
Features: (b) property filter, (c) guided search, (d) autocompletion suggestions, (e) number of search results, (f) selection shortcut, (g) search results, (h) visualization selection options, (i) Scatterplot 2 properties selection, (m) Table set selection.
(j1) The scatterplot shows the distribution of Van Gogh's self portraits and Rembrandt's self portraits with respect to the date and material used. The visualization shows that Rembrandt had consistently painted a small number of self portraits distributed throughout many years using different kinds of materials (e.g. canvas, Oak panel, Mahagony panel). Van Gogh, however, made many self-portraits between 1886 and 1887. In 1887 alone, he made about 11 oilpaint (olieverf) self-portraits.
(j2) The table shows all values of Van Gogh's self-portraits. The smallest height of a van Gogh's self-portrait is 19 cm (n).
(j3) The bar chart shows the distribution of materials used for Set A and Set B. 20 of Rembrandt's self portraits are made out of oilpaint (olieverf) and 13 are made out of Oak (Eikenhout).

(AAT), the Getty Union List of Artist Names (ULAN) and the Getty Thesaurus of Geographical Names (TGN) [9], as well as collection specific thesauri, such as thesauri from RKD. Table **??** shows the size of collections and thesauri currently used by the application. Collections and thesauri data were converted to an RDF/SKOS representation. To allow information access across collections, specific thesauri are aligned with the common ones. For example, artists' names in the RKD thesauri are linked to artists' names in ULAN. Materials terms in the RKD thesauri are linked with concepts in AAT. Detailed information on the conversion and alignment methods of cultural heritage sources used can be found in (Tordai et al. 2007; van Assem et al. 2006).

Table 5.2: Thesauri and collections used in LISA*)

source (thesaurus coverage)	**size**	
Collection:		
RKD Archive	82.781	objects
Thesaurus:		
RKD thesaurus (RKD)	11.995	terms
TGN (geographical)	89.000	terms
ULAN (artist)	13.000	people
AAT (art and architecture)	31.000	terms
IconClass (iconographic)	24.331	terms

5.6 Evaluation Study: Thesaurus-based comparison search interface evaluation

The goal of this study was to evaluate LISA. In particular, we focus on evaluating how well the search, selection and comparison features support experts' comparison search tasks. As a baseline, we use the RKDimages website (`http://english.rkd.nl/Databases/RKDimages`), a popular online cultural heritage archive that contains descriptions, metadata and images of Dutch and Flemish artworks from the 14th-19th century. The RKDimage website contain a comprehensive coverage of different artworks and is widely used as reference. For the purpose of the evaluation, participants can access the same information with LISA as well as with RKDimages website. The research questions for this evaluation are:

- Does the LISA tool support searching, selecting and comparing artworks more

[9] `http://www.getty.edu/research/conducting_research/vocabularies/`

efficiently than the baseline tool?
• Does user perceive the LISA tool asier to use than the baseline tool?

5.6.1 Setup

The experiments took place at the participants working place. Every participants was asked to complete comparison search tasks with LISA and RKDimages (within subject design). The experiment had four parts:
• Introduction. Participants were asked demographic questions and for informed consent. Afterwards, they were shown a video demo and were asked to perform trial tasks on LISA. As most participants were regular users of the RKD website already, the demo focused on familiarizing the participants with LISA.
• Experiment session. The experiment was divided into two phases. In the first phase, we compared how well the 4 different type of presentations (LISA Table, LISA Bar Chart, LISA Scatterplot, RKDimages) support single property comparison. In the second phase, we compared how well 3 different type of presentations (LISA Table, LISA Scatterplot, RKDimages) support dual property comparison [10]. In total, participants were given 14 comparison tasks (2 tasks per interface). At the completion of each task, participants were ask about how they perceived the ease of use of the interface used. At the end of all task, participants gave general impressions about LISA and RKDimages.

The user recruitment were based on email invitations and open invitations on a cultural heritage online forum. In total there were 12 cultural heritage experts from seven cultural heritage institutions (see Table 5.2). Our participants conduct searches within collections frequently for their work (M=4.8, SD= 1.8) [11] and are fairly familiar with the RKDimages website (M=3.8, SD= 2.5) [12]. None of the participants have used the LISA interface prior to this evaluation.

[10] The LISA Bar chart was omitted in the second phase because it is unsuitable for dual property comparison
[11] Seven point scale, 1: not very often, 7: very often
[12] Seven point scale, 1: very unfamiliar, 7: very familiar

Table 5.3: User Study 2: Participants demography (total: 12 people)

Age:	21-60 years old (M=39.6, SD=12.1)
Gender:	4 male, 8 female
Affiliation	CH institution(6), museum(2),
	Art historical archive(3), university(1)
Expert role:	researcher(4), curator(2), ICT(2)
	program manager(3), librarian(1)

Table 5.4: Searching and Selecting artworks with LISA and RKDimages

	LISA	RKDimages
a. Search and select avg. time in min (SD)		
1. FEW	1.38(.41)	1.58(1.19)
2. MANY	1.37(.51)	3.33(1.90)
b. Search - ease of use score (SD)		
1. FEW	5.75(1.22)	5.58(1.38)
2. MANY	6.08(1.08)	4.67(1.78)
c. Select - ease of use score (SD)*		
1. FEW	5.92(1.16)	4.50(1.68)
2. MANY	6.00(0.95)	3.58(1.68)

* 7-Likert point scale, score 1:very difficult, 7:very easy

Few: few objects comparison (1 item/set)

Many: many objects comparison (10-15 items/set)

5.6.2 Task

We use two independent variables as dimensions that reflects the complexity of comparison search.
• Amount of artworks. We have two conditions: comparing *few* objects (1 artworks per set) and comparing *many* artworks (10-15 artworks per set). We expect that the more artworks there are, the harder it is to perform comparison search.
• Number of property. We have to conditions: *single* property comparison and *dual* property comparison. An example of a comparison search task is as follow: *Use the scatterplot to answer this question: Compare all artworks having the subject depiction church from the Stedelijk Museum De Lakenhal with all artworks having*

Table 5.5: Comparing artworks with LISA and RKDimages

a. Single property comparison avg. time in min. (SD)				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	0.91(.75)	0.94(.34)	1.32(.66)	0.83(0.24)
2. MANY	1.37(.55)	1.19(.36)	1.70(.76)	2.13(1.61)
b. Dual property comparison avg. time in min. (SD)				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	1.14(.71)	-	1.28(.63)	1.06(.40)
2. MANY	2.28(.97)	-	1.39(.65)	2.99(.99)
c. Compare - ease of use score (SD) *				
	Table	Bar chart	Scatterplot	RKDimages
1. FEW	4.92(1.68)	5.42(1.08)	4.50(1.78)	2.83(1.47)
2. MANY	5.08(1.38)	4.75(1.54)	4.75(1.76)	2.17(1.47)

* 7-Likert point scale, score 1:very difficult, 7:very easy

the subject depiction church from Museum Bredius.
(1) Which artist made the most artwork?
(2) How many of these artworks are painting and are made after 1612?

Example 1 is a single property comparison task for the property `Artist`. Example 2 is a dual property comparison task for the properties `Object type` and `Date`.

5.6.3 Results

In this section we discuss observations on how well LISA and the RKDimages website support the comparison search tasks conducted by the participants.

Searching and selecting artworks — Prior to the study, we had an assumption that search and selection of artworks are two separate and independent activities. However, during our experiments we observe this is not the case. Participants conduct search, make selections, go back to repeat search before being satisfied with the selection result. Thus, we will combine our observations for this two activities.

We analyzed the time performance and the perceived ease of use for searching and selecting artworks. Table 5.3a. shows the average time it took for participants to complete searching and selecting for (Few and Many) artworks with LISA and RKDimages website. Using the Wilcoxon Sign-rank (WSR) test[13], we found that participants were about two times slower when searching and selecting many artworks using RKDimages (*Mdn*=2.75 min) than searching and selection few artworks (*Mdn*=1.12 min) z=-2.04, p<.05, r=-.42. This confirms with our expectation that the more artworks there are, the more time it will take to search them using the baseline tool. When participants use LISA for searching and selecting artworks, however, the results were different. We did not found significant difference when participants were searching and selecting for few artworks (*Mdn*=1.43 min) and for many artworks (*Mdn*=1.15 min). Moreover, when searching and selecting many artworks, RKDimages (*Mdn*= 2.75 min) is significantly slower than LISA (*Mdn*= 1.15 min) z=-2.43, p<.05, r=-.49. Thus, we conclude that LISA is more time efficient than RKDimages for searching and selecting many artworks.

Participants also rated the perceived ease of use of LISA and RKDimages for searching and selecting artworks (see Table 5.3b,c). RKDimages is significantly easier to use when searching for few artworks (*Mdn*=6.0) than when searching for many artworks (*Mdn*=5.0) z=-2.46, p<.05, r=-.50. When searching with LISA, however, this difference was not significant. Searching for many artworks (*Mdn*=6.0) is perceived as easy as searching for few artworks (*Mdn*=6.0). Participants also think that RKDimages (*Mdn*=5.0) is harder to use than LISA (*Mdn*=6.0) when they need to search for many artworks z=-2.38, p<.05, r=-.48. Thus, we conclude that LISA is easier to use than RKDimages specially for searching many artwork. We asked the same question for selecting artworks and found congruent results. We found LISA significantly easier to use than RKDimages when selecting few artworks as well as many artworks (z=-2.06, p<.05, r=-.42 and z=-2.69, p<.05, r=-.55. respectively). Based on this, we conclude that LISA is easier to use than RKDimages with respect to selecting few as well as many artworks.

For searching and selecting artworks, we found LISA more time efficient and easier to use than RKDimages. We also found that unlike the time efficiency with the baseline tool, time efficiency with LISA does not suffer much as there are more artworks to search and select. Moreover, overall, participants perceived LISA easier to use than RKDimages for searching and selecting artworks.

Comparing artworks — We use Friedman Analysis of Variance by Ranks (FAVR) to examine if there were any difference in time performance between the different presentations types for single property comparison and dual property comparison tasks (see Table 5.4).

[13]Nonparametric statistics is used throughout the study as not every data meet parametric assumptions.

• Single property comparison search tasks
For single property comparison, we compared the participants time efficiency using four different presentations: LISA-Table, LISA-Bar chart, LISA-Scatterplot and RKDimages. Using the FAVR test, we found no significant difference for the four different presentations. This applies for single property comparison task for few artworks ($\chi^2(3)$=3.7, p>.05) as well as for many artworks ($\chi^2(3)$=3.0, p>.05) We conclude that all four presentations performs equally with respect to the time spent to conduct single property comparison.
• Dual property comparison search tasks
In dual property comparisons, we found no significant difference between the three presentation for few artworks comparisons ($\chi^2(2)$=2.2, p>.05). However, there is a significant difference between three different presentations (i.e. LISA-Table, LISA-Scatterplot and RKDimages) for many artworks ($\chi^2(2)$=9.5, p<.05). We found significance of of time performance between all three presentations. LISA-scatterplot begin the fastest (*Mdn*=1.26 min), followed by LISA-table (*Mdn*=2.08 min), and RKD images (*Mdn*=3.02 min) is the slowest [14].

We saw clearer trends for the perceived ease of use scores (see Table 5.4c). We found significant differences for comparison search tasks in few artworks ($\chi^2(3)$=13.07, p<.05) as well as many artworks ($\chi^2(3)$=17.49, p<.05). The WSR post-hoc test confirms that All LISA visualization is perceived easier to use than RKDimages for comparison. Thus, we conclude that comparing with LISA is more time efficient than with RKDimages mainly for many artworks dual property comparison seach task. However, participants perceive LISA easier to use that RKDimages for few as well as many artworks comparison.

5.7 Discussion

The thesaurus-based comparison search tool builds on two important components: the interface and the data. We discuss challenges and improvements with respect to these.

5.7.1 Interface

Searching and selecting artworks — The evaluation study showed that searching and selecting artworks with LISA gives better time performance than with RKDimages. This is mainly because of the ease of use of the thesaurus-based

[14]WSR post-hoc tests shows LISA Scatterplot - LISA Table (z=-2.12, p<.05, r=-.43), RKDimages - LISA Table (z=-1.96, p<.05, r=-.40), RKDimages - LISA Scatterplot (z=-2.83, p<<.05, r=-.58),

guided search and the selection interface. The thesaurus-based autocompletion enable users to quickly find the correct term to find the artwork. This aligns with previous research on the use of thesaurus-based autocompletion for term search (Amin et al. 2009). To increase confidence levels when selecting terms from the autocompletion suggestions, improvements, such as adding extra information about the terms, e.g. by showing the thesauri hierarchy or descriptions of the terms, could be made (Hildebrand et al. 2009).
Comparing artworks — We found that, with respect to the time spent, LISA shows little improvement on RKDimages for comparison activities. There can be several explanations for this. Our participants are regular RKDimage users, thus they are more experienced in using this application for comparison. Most of them, however, are not used to handling graphs and charts. We observed that participants need to spend time to become familiar with the tool. We acknowledge that we cannot fully eliminate learning effects in a one time evaluation study of a complex tool such as LISA. Many of the participants thought they would be able to handle LISA better once they were accustomed to it. *"I think you have to get used to the system, like to any new system"* [P1], *"This is a new way to present and interact with (museum) collections."* [P14].

Even though LISA did not significantly improve the speed of comparison, participants clearly favor LISA above RKDimages with respect to ease of use. They see the practical benefit of having aggregated results presented automatically rather than computed manually. Before trusting the results, however, they need to understand how the thesaurus-based aggregation works to produce the graphs and charts of the presentations. Participants also appreciated the different presentations as they provide more ways to analyze the same data. *"I think there are different learning and reading styles, so it is useful to have these variations"* [P3].
Additional features — Sometimes experts need to go back and re-examine previous comparison search tasks. Experts may also want to save the visualization results and include them in a report, or may need to inspect the visualization in detail. Our experts listed features such as bookmarking, search history and 'save as' as additional functionalities that LISA should have. Participants also mentioned two important visualization improvements. First, the ability to enlarge the visualization size on demand. This feature is specially useful when dealing with many artworks. Second, more interactivity with the visualization, such as being able to trace back from the visualization to the original artworks.

5.7.2 Data

Based on what we have learned during the LISA implementation and the evaluation study, we identify characteristics of the collection metadata that developers need be aware of when developing thesaurus-based aggregator services.

Semantic aggregation — Thesaurus-based comparison search should take into account semantic aggregation where narrower/broader relationships exist between terms. For example, to be able to answer the comparison search question "How many artworks are paintings?" correctly, the system needs to quantitatively aggregate not only all artworks having `object type: painting`, but also all artworks annotated by the narrower concepts of painting, such as `object type: aquarel`, since Aquarel is a type of a painting.

Inconsistent data — Museum collection metadata may be inconsistent, for example in measurement units. Artwork dimensions, such as height and width, can be specified in different units, e.g. feet, cm or mm. Prior to an aggregation process, metadata needs to be cleaned.

Incomplete metadata — In reality, museum metadata is not always complete. Parts of the collection may have insufficient or missing values. A quantitative aggregation on these data will generate false results. One solution is to check and improve the quality of data automatically as suggested in (van den Bosch et al. 2009).

Estimated data — In some cases, the metadata contains an estimated value, e.g., the creation date of an artwork is simply unknown. The museum is able to supply only an estimate, e.g. "before AD 400" or "between 600-700". Providing accurate aggregation results based on these estimated data is not possible.

Quality of data alignment — The accuracy of information presented in a thesaurus-based aggregated system also depends on the quality of the data alignment, i.e. linking metadata to terms from individual thesauri and linking terms among different thesauri. Methods for vocabulary alignment are still at a preliminary stage of development, e.g. (Tordai et al. 2009), making it difficult to predict when automatic methods will be of sufficient quality for our experts' needs.

The biggest opportunity for LISA lies in the area of comparison across different collections. At the moment, such tasks are carried out by accessing the different collections individually, and then integrating the data manually. While there is room for improvement by being able to automatically aggregate over multiple collections, the number of errors in the results will also grow when performing computations across collections because of the different schemata and thesauri used to describe the artworks. When dealing with real museum datasets, we believe that the answer needs to come from making the computations more transparant in the interface, allowing the user to examine the data and how the computation is being performed. Thus, if there are errors in the results, the user should be able to trace them, make corrections and even correct the underlying data.

5.8 Conclusions and Future work

We conducted a user-centered design study on a comparison search application for the cultural heritage domain. In a preliminary study, we identified various comparison search use cases, such as learning about collections, planning exhibitions, museometry and qualitative comparison, and identified the challenges users face while performing comparison search, such as searching and selecting terms, and comparisons involving multiple sets and properties. In our evaluation study, we found that our comparison search tool can help users, in particular for efficient searching for terms and selecting artworks. In general, participants perceived the comparison search tool as easier to use than the baseline tool with respect to searching, selecting and comparing artworks. Finally, based on our implementation experience and evaluation study, we identify future improvements to the interface, namely, supporting interactivity visualizations, improving the autocompletion and providing bookmarking and search history functionality. For the data issues, we would be able to extend LISA's functionality to support semantic aggregation. For the other data issues we are dependent on others making the data more reliable. Making the issues more transparent through the LISA interface would be one way of tackling this.

CHAPTER 6

The Effects of Source Credibility Ratings in a Cultural Heritage Information Aggregator

One important aspect of information seeking in multiple sources is finding and selecting reliable information sources. We investigate the effect of displaying credibility ratings of multiple sources for the cultural heritage domain. Our aim is to investigate whether lay users' experiences while accessing information can be enhanced through visualization of credibility ratings (research question 3). In the first of two studies, we investigated whether there is a difference in how lay users perceive the credibility of different types of cultural heritage information sources, such as museums, news organizations and blogs. In the second study, we investigated whether source credibility has an influence on lay users' confidence when they are accessing information. The results of our online interactive study show that by presenting the source credibility information explicitly, people's confidence in their selection of information increases, even though it does not necessarily make search more time efficient. We also identify credibility issues that are applicable beyond the cultural heritage domain, such as issues related to credibility measures and choice of visualization.

This chapter was published as "The Effects of Source Credibility Ratings in a Cultural Heritage Information Aggregator" in the Workshop on Credibility on the Web (WICOW'09) at WWW'09, (Amin et al. 2009) and as a poster titled "Improvings user's confidence in cultural heritage aggregated results" in the Special Interest Group in Information Retrieval conference (SIGIR'09) (Zhang et al. 2009) and was co-authored by Junte Zhang, Lynda Hardman, Henriette Cramer and Vanessa Evers

6.1 Introduction

Consider a student who is looking for information about Dutch painters in the 17th century who made portraits in a certain artstyle. This information may be online, but bits and pieces may exist on different HTML web pages, Wikipedia articles, weblogs, etc. The student has to choose which pages to look at from a large number presented in the search result page. Before making this decision, s/he may examine where the information comes from and who wrote it before looking at the article itself.

The situation above occurs frequently to all of us in slightly different settings. Many web search aggregators enable us to find information from different sources simultaneously (Murdock and Lalmas 2008). These systems enable us to quickly retrieve information from multiple sources but the decision as to which information source to go for is something that is left to the user. When interacting with such systems, we constantly need to assess the information sources, the authors and the content. In this decision process, credibility plays an important role.

Even though the number of information aggregators is expanding, there are only few studies that report on how the added complexity of having to deal with multiple information sources influences the users' ability to make decisions and select the appropriate information. The aim of this study is to investigate how displaying the credibility ratings of cultural heritage sources affects the user's confidence and time to search for information.

The outline of this chapter is as follow: related work and problem statement are given in the next sections, the experimental setup is laid out in section 6.4, the results are presented in section 6.5 and finally we end this chapter with some conclusions and future work in section 6.6.

6.2 Related Work

6.2.1 Credibility and the Web

We adopt the simple notion of credibility as *believabilty* (Fogg and Tseng 1999). One of the earliest research about credibility was conducted by (Hovland, Carl I. and Weiss, Walter 1951). That paper specifically focused on *source credibility*, i.e. credibility of a source. More recently, it is becoming an important issue for research on the World Wide Web and information access as well. As new information systems are emerging that combine information from multiple sources, the effects of source credibility for improving information access comes back as a research issue. In (Flanagin and Metzger 2007), distinctions are made between several types of credibilities, like Web credibility, site credibility, sponsor credibility, news credibility, etc. Credibility research is complex, and researchers do not necessarily or completely agree with its dimensions as a concept. However, a comprehensive and extensive literature overview about credibility is presented in (Rieh and Danielson 2007), where credibility is related to a multidisciplinary framework, and related to other concepts (which are not the same as credibility), such as 'quality', 'authority', 'trust', and 'persuasion'. There are also different types of credibility by attaching it to potentially interlinked objects of assessment, like source credibility, media credibility, and message credibility. In this chapter we mainly focus credibility on the object of the source.

There has been ample research conducted about credibility, and in a plethora of domains. For example, the credibility in the online encyclopaedia Wikipedia (Chesney 2006; Lopes and Carriço 2008), and website design for achieving credible websites (Tseng and Fogg 1999). Credibility in academic information and in the online encyclopaedia Wikipedia (Metzger et al. 2003; Chesney 2006), health and medicines (Lindberg and Humphreys 1998; Eastin 2001; Walther et al. 2004; Craigie et al. 2002; Peterson et al. 2003; Eysenbach and Kohler 2002), media and news providers (Kiousis 2001; Sundar and Nass 2001; Tsfati and Cappella 2005; Choi et al. 2006; Cassidy 2007), website design for achieving credible websites (Tseng and Fogg 1999; Hong 2006; Flanagin and Metzger 2007). These research reports different aspects of credibility. For example, in (Eysenbach and Kohler 2002) the objectives were to describe techniques for retrieval and appraisal used by consumers when they search for health information on the Internet. They identified factors to determine the website credibility, where authority of a source, email, credentials and qualifications can be applied to the source credibility. The credibility of the source appeared to be a common determinant in the criteria of all participants who looked online for information about medicines (Peterson et al. 2003). Source credibility is also a very important issue in the cultural heritage domain, especially for cultural heritage experts (Amin et al. 2008) and historians (Duff et al. 2004). Both studies have reported that it is important for these experts to be able to assess the credibility of the source before using their information.

6.2.2 Aggregated Search

As defined in (Murdock and Lalmas 2008), Aggregated Search deals with the task of searching and assembling information from a variety of sources, and placing that information in a single interface. Examples of generic Information Retrieval (IR) systems are Alpha Yahoo![1], Google Universal Search[2] and Naver[3], a Korean search

[1] http://au.alpha.yahoo.com/
[2] http://www.google.com/intl/en/press/pressrel/universalsearch_20070516.html
[3] http://www.naver.com/

portal. There are also domain specific IR systems, such as Google Scholar[4], a search engine that harvests information from publishers, preprint repositories, universities and scholarly organizations, or WorldCat[5], a union catalog of more than 10.000 libraries. In the cultural heritage domain, ECulture Multimedian(Schreiber et al. 2006) and CultureSampo(Hyvonen et al. 2009) allow to search for information from various museums or similarly the README system (Zhang et al. 2008) for historical archives.

These systems were typically aimed at (semi-)expert users, such as hobbists, scholars or professionals. For these systems, several issues remain open questions, such as how to harvest and to present high quality information (Murdock and Lalmas 2008), how to tackle information conflict or different views among the information sources, and how to present the credibility of information sources.

6.2.3 Transparency

In the context of the evaluation of IR systems, it has been pointed out that a central question for the design of interactive systems is the amount of knowledge that is needed about a system (Koenemann and Belkin 1996). On the one hand, interfaces can hide the inner workings of a system as much as possible and put the focus on the user's task. However, on the other hand, some knowledge and control may be necessary to enhance interaction with (different components of) the system. This deals at its core with the issue of *transparency*. Conversely, transparency could also influence the trust and acceptance of systems.

As pointed out in (Cramer et al. 2008), there has not been a clear-cut consensus in numerous different studies which empirically tested the effects of system transparency. There were mixed findings: transparency could enhance or even worsen user interaction with systems. There is no general guideline and it highly depends on the implementation and the application. For example, the study that was conducted in (Cramer et al. 2008) tested the effects of transparency with an adaptive recommender system. It was found that transparency increased the acceptance of recommendations and makes a system more understandable, which correlates with the perceived competence of users (and thus enhances the interaction with the system).

6.2.4 Credibility Measures

In a computer credibility research (Tseng and Fogg 1999), credibility is interpreted as believability. The authors point out that credibility has 2 dimensions: The *trust-*

[4] http://scholar.google.com/
[5] http://www.worldcat.org/

worthiness dimension of credibility captures the perceived goodness, morality of the source or 'well-intentioned', 'truthful', and 'unbiased'. The *expertise* dimension of credibility captures the perceived knowledge and skill of the source, or 'knowledgeable', 'experienced', and 'competent'. They continue by pointing out that credible people are believable people; credible information is believable information. Credibility is a perceived quality, reflected by phrases, such as 'trust the information', 'accept the advice', 'believe the output'.

Most research on credibility take quantitative empirical social science experiments as their approach. In this chapter we adopt a similar approach.

6.3 Problem Statement

When presented with multiple sources users need to make several decisions about the information search results. This added complexity is likely to influence users' search performance. Our main research question is therefore:
Does source credibility improve information access to information aggregated from multiple sources?

We elaborate our research question through three research hypotheses. First, when the users are confronted with multiple sources, they need to establish the credibility of the source and its information. We expect that the users will feel less confident in their selection of information from the numerous aggregated results:

H1 *The confidence users have in the reliability of the information they selected will decrease when the number of sources increases.*

Results from (Cramer et al. 2008) indicate that information transparency can positively influence the acceptance of recommendations coming from a single source. This may also be the case for information from multiple sources provided by an information aggregator. By presenting the origin of the information and the credibility ratings, a user may be able to assess information quicker and be more assured about their decisions. Our conjecture is that it will result in higher confidence levels and less time needed to search. This leads to the following hypotheses:

H2 *Displaying the 'source credibility' will give users greater confidence in the information they select.*

H3 *Displaying the 'source credibility' will produce a shorter search time when compared to when it is not displayed.*

6.4 Experimental Setup

To test our hypotheses, we conducted a two phase studies. The first study was a pre-study designed to elicit source credibility scores for the cultural heritage sources, such as museums and art websites, that will be used for the Study 2. The second study aimed to investigate how displaying source credibility influences the user search efficiency.

6.4.1 Study 1: Measuring Cultural Heritage Source Credibility

For the purpose of eliciting credibility scores of 12 art sources from 4 categories (see Table 6.1) we developed an interactive online survey (see Figure 6.1) which automatically generated a random set of 6 art sources, out of a total of 12, for the participant to assess on credibility.

6.4.1.1 Procedure

When evaluating credibility of an online information, both the receptor's attribute (Berlo et al. 1970) and the source's attribute (Tseng and Fogg 1999; Flanagin and Metzger 2007; Hong 2006) should be taken into consideration. Based on this literature, we define the credibility measures for the cultural heritage domain:

Receptor's attributes knowledge (of arts and culture), and reliance (degree of importance of arts and culture).

Source attributes Trustworthiness, Completeness, Goodwill, and Expertise.

Participants were asked demographic questions reflecting the receptor's attribute. To measure the source attributes, participants are given information about the different sources taken from the sources' public website. The type of chosen information were taken from(Eysenbach and Kohler 2002; Metzger et al. 2003) and adapted for the cultural heritage domain. This information is: *logo*, *name*, *contact information*, *mission statement*, *history*, *repository*, *source type*, and *organization.* For each source, participants were asked to assess the 4 source attributes by using a 5-point scale (see Fig. 6.1).

Trustworthiness (T) *"I believe this source will give cultural information that is neutral and it has good intentions."*

Completeness (C) *"I believe this source is able to give me everything I need to know about an artwork."*

Goodwill (G) *"I believe this source genuinely cares about arts and culture."*

Expertise (E) *"This source appears to be a leader in its area of speciality."*

Additionally, participants were also asked whether they were familiar with or have prior knowledge about each source by using 5-point Likert scale.

Source knowledge *"I already have a lot of knowledge about this source."*

6.4.1.2 Participants

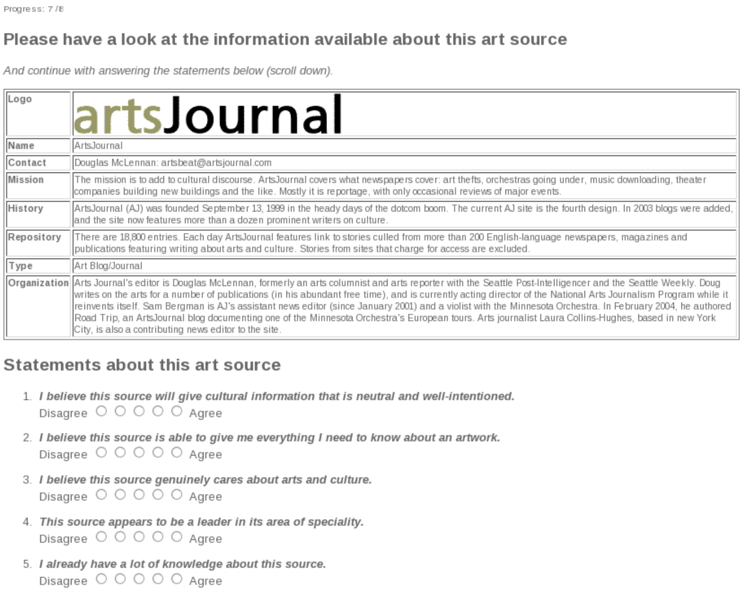
In total, 57 participants were recruited from Dutch online discussion boards over a period of 3 weeks. The gender is almost evenly distributed (male: 47.4%, female: 52.6%), they were highly educated (61.4% University education), were very experienced with the Internet ($M = 4.39$, $SD = 0.80$), had some arts and culture knowledge ($M = 3.14$, $SD = 0.93$), and had lived their entire life in the Netherlands ($M = 25.49$ years, $SD = 9.93$).

6.4.1.3 Cultural heritage source credibility scores

To examine the reliability of our 4 source attributes value, we calculated the Cronbach's α and found the value of 0.70, which is sufficient to determine the credibility of a source.

Source	Credibility score M (SD)	γ	Source Type
Joods Historisch Museum	4.09 (.61)	6.7	
Rijksmuseum A'dam	4.34 (.59)	7.5	T1
Museum Volkenkunde	4.05 (.55)	6.6	
absolutearts.com	3.75 (.55)	5.6	
ArtLex Art Dictionary	3.25 (.77)	4.2	T2
Grove Art Online	3.96 (.69)	6.3	
art.blogging.la	3.20 (.66)	4.1	
Artblog.net	3.13 (.74)	3.9	T3
ArtsJournal	3.29 (.66)	4.3	
About.com	2.94 (.82)	3.5	
Infoplease	2.93 (.72)	3.4	T4
Wikipedia	3.05 (.86)	3.7	

Table 6.1: Sources' Credibility Score, where T1: Museums, T2: Arts Websites, T3: Art Blogs, T4: General Websites. γ represents the marked up credibility score as calculated by equation (**??**) and (**??**).


Progress: 7 /8

Please have a look at the information available about this art source

And continue with answering the statements below (scroll down).

Logo	artsJournal
Name	ArtsJournal
Contact	Douglas McLennan: artsbeat@artsjournal.com
Mission	The mission is to add to cultural discourse. ArtsJournal covers what newspapers cover: art thefts, orchestras going under, music downloading, theater companies building new buildings and the like. Mostly it is reportage, with only occasional reviews of major events.
History	ArtsJournal (AJ) was founded September 13, 1999 in the heady days of the dotcom boom. The current AJ site is the fourth design. In 2003 blogs were added, and the site now features more than a dozen prominent writers on culture.
Repository	There are 18,800 entries. Each day ArtsJournal features link to stories culled from more than 200 English-language newspapers, magazines and publications featuring writing about arts and culture. Stories from sites that charge for access are excluded.
Type	Art Blog/Journal
Organization	Arts Journal's editor is Douglas McLennan, formerly an arts columnist and arts reporter with the Seattle Post-Intelligencer and the Seattle Weekly. Doug writes on the arts for a number of publications (in his abundant free time), and is currently acting director of the National Arts Journalism Program while it reinvents itself. Sam Bergman is AJ's assistant news editor (since January 2001) and a violist with the Minnesota Orchestra. In February 2004, he authored Road Trip, an ArtsJournal blog documenting one of the Minnesota Orchestra's European tours. Arts journalist Laura Collins-Hughes, based in new York City, is also a contributing news editor to the site.

Statements about this art source

1. ***I believe this source will give cultural information that is neutral and well-intentioned.***
   Disagree ○ ○ ○ ○ ○ Agree
2. ***I believe this source is able to give me everything I need to know about an artwork.***
   Disagree ○ ○ ○ ○ ○ Agree
3. ***I believe this source genuinely cares about arts and culture.***
   Disagree ○ ○ ○ ○ ○ Agree
4. ***This source appears to be a leader in its area of speciality.***
   Disagree ○ ○ ○ ○ ○ Agree
5. ***I already have a lot of knowledge about this source.***
   Disagree ○ ○ ○ ○ ○ Agree

Figure 6.1: Example screenshot used in Study 1 where participants judge the credibiltiy of a cultural heritage source

The credibility scores of the 12 sources are depicted in Table 6.1. The scores of source attributes were averaged into a single numeric value. The museums were perceived as most credible by the participants, followed by arts websites. The general websites, which do not exclusively cover arts and culture, score lowest. Interestingly, *Groove Art Online* was considered almost as credible as the museum sources. Overall, the sources were assessed as what we expected beforehand. We also checked for the relationship between source knowledge and the perceived credibility score. We found weak relationship between source knowledge and source credibility. Only *Art.blogging.la* was significant (2-tailed) (r=.39, n=28, p<.041) and no correlation for the other sources. We conclude that for the sources that were used in this study, participants gave the credibility scores based only on the type of information given (e.g. source history, mission statement) and independent of their knowledge with the source.

We use these credibility scores as basis for our follow up study. To increase the difference between the sources, we mapped the credibility score by using a quadratic function (see eq. **??**). This results in larger differentials between the scores which will make the visualization clearer for Study 2[6]. The score Ψ was calculated for each score by taking the quadratic value of credibility scores and translated to a 0 to 10 scale (0: lowest credibility, 10: highest credibility).

$$\Psi_i = \left(\frac{T_i + C_i + G_i + E_i}{4}\right)^2 \cdot \frac{1}{2.5} \tag{6.1}$$

$$\gamma = \frac{1}{K_{total}} \cdot \sum_{i=1}^{N} (\Psi_i \cdot K_i) \tag{6.2}$$

where the rating T is the *Trustworthiness*, rating C is *Completeness*, rating G is *Goodwill*, rating E is *Expertise* for the i-th participant, and $\Psi \in [0, 10]$. Finally, we take the users source knowledge into account for the mark up credibility score of each source γ by taking Ψ from each i-th participant, and use the *Source Knowledge* K_i as propagation factor, where $K_i \in [1, 5]$ and K_{total} is the sum of all K-s (see Table 6.1 for mark up credibility score γ for all sources).

6.4.2 Study 2: Cultural Heritage Source Credibility Effects

In Study 2 we investigated the effects of displaying credibility scores of cultural heritage sources on users confidence and search time. We use the mark up credibility scores γ from Study 1 in a bar visualization for different cultural heritage information sources.

6.4.2.1 Procedure

An experiment was carried out to assess the effects of multiple information sources and source credibility on search performance and user confidence. A 2x2 experiment was carried out with number of sources (few = 4 sources, many = 12 sources) and source credibility (ratings or no ratings) as between-subject variables (see Table **??**). An interactive online survey was developed in which participants were randomly and automatically assigned to one of the four conditions (FN, FR, MN, or MR).

Each participant was given 3 tasks. In each task, a participant needs to select the culture of origin and the description of artwork that seemed most accurate to

[6]Alternatively, this mapping could be done with other functions, such as logarithmic.

	Few Sources (4 sources)	**Many Sources (12 sources)**
No credibility rating	*FN*	*MN*
With credibility rating	*FR*	*MR*

Table 6.2: Study 2 setup with 4 conditions: the number of cultural heritage information sources (few or many) vs. availability of the source credibility score ratings (displayed or not displayed).

them. The possible answers shown come from different cultural heritage information sources. In the FN and MN conditions, the participant only sees the name of the information source. In the FR and MR condition, the participant is also given extra information on the credibility ratings of each information source (see example in Fig. 6.2 from Study 1). In total, participants conducted 3 tasks. The tasks order is randomized across all participants. After each task, participants were asked to rate the confidence of their selection in a 5-point scale.

6.4.2.2 Variables

The descriptions available for the artwork were all designed to be concise, equally realistic and plausible for all four conditions. Number of sources was manipulated by either presenting information from 4 or 12 sources. As a greater number of choices also implies a longer time needed to complete a choice task (Duff et al. 2004), this effect were taken into account. The answers for the many sources conditions were short sentence, for the few sources conditions answers were usually longer (around 3 sentences).

Source credibility transparency was manipulated by displaying (or hiding) the source credibility for each chunk of information from a different source. Each task consisted of low and high credibility sources. In the no source credibilility conditions (FN and MN) the orange rating bar and the credibility indication were omitted (see Figure 6.2). We presented low and high credibility ratings (not controlled), because we wanted to simulate a more realistic use case and interactive scenario.

We wanted to focus on and manipulate only the influence of credibility of source and not the credibility of content. Thus, in every task, the information content provided for the artwork was fictional but equally plausible (as assessed by multiple art experts). We have also chosen unknown artworks to make sure that the participants could not rely on prior knowledge about the artwork to answer the questions in the task.

Figure 6.2: Example screenshots used in Study 2. *Top*: few sources-with credibility ratings condition (FR). *Bottom*: few sources-no credibility ratings condition (FN).

The dependent variables measured were *efficiency* (the time it took a participant to finish the search task), and participant's *confidence* in the accuracy of the information selected. *Search accuracy* (giving the right answer) was not measured as the information given to the participants were fictional but equally plausible.

We also checked whether participants rely on the content of the information or the source (Q1 and Q2). Additionally, for the rating conditions (FR and MR), we asked an extra question on the visualization (Q3). The answer selected by the participant was expressed on a 5-point scale.

Q1 Reliance on source *"I only choose the answer from the source I trust most."*

Q2 Reliance on content *"I only choose the answer that seems most correct to me."*

Q3 Reliance on visualization *"I only choose the answer based on the visualization."*

6.4.2.3 Participants

Participants (N = 122) for Study 2 were recruited in the Netherlands through a mailing to students and research staff at Dutch research institutes in a period of 2 weeks. The valid response ratios were 81.6% for FN (N = 30), 71.4% for MN (N = 31), 81.6% for FR (N = 30), and 75.0% for MR (N = 31).

The participants demographics were similar to those in the first study. The gender was equally divided, the average age was 29 years old (SD=10.15), they were highly educated (68.0% University level), were very experienced with the Internet (M=4.50, SD=.71), had modest knowledge of Cultural Heritage (M=2.71, SD=.92), but they gave a high importance to Cultural Heritage (M=3.64, SD=1.00) and most had lived their entire lives in the Netherlands (M=25.07 years, SD=12.35).

6.5 Results

In the analysis, Kruskal-Wallis (K-W) test was performed to check for the effects across the four groups[7]. Afterwards, the Mann-Whitney (M-W) U test was used as a post-hoc test to check for effects on the dependent variables (efficiency, confidence, reliance of content and reliance on source). We conducted the M-W U test for both the few sources conditions (FN and FR) and the many sources conditions (MN and MR), and for with credibility ratings (FN and MN) and no credibility ratings conditions (FR and MR).

6.5.1 Effects of Displaying Credibility Ratings

A significant effect was found on confidence in the accuracy using the K-W test (χ^2(3)=11.16, p(2- tailed)=.011). Scores on confidence were highest for both with credibility ratings/many sources - MR (M=3.90) and the with credibility ratings/few sources - FR (M=4.19) conditions. Therefore we accept Hypothesis 2: displaying the 'source credibility' ratings do give users greater confidence in the information they select.

There was also a significant effect on efficiency using the K-W test (χ^2(3)=17.80, p(2-tailed)=.000). The efficiency was surprisingly the highest for the no credibility rating/few sources condition (FN), where the average rank was 47.16 (M=214.26 s) and the non-transparent/many sources condition with an average rank of 69.63 (M=458.03 s). This means we cannot support Hypothesis 3.

The M-W U test was used to check for more effects. In the few sources condition, the effect of transparency on reliance on content was significant (U = 341.500,

[7] Nonparametric statistics is used as the data did not meet parametric assumptions

z = -2.1, p(2-tailed)=.038). When the source credibility rating is displayed, the average rank was 27.02 (Mdn=4.00), whereas when it was not displayed, there was an average rank of 35.98 (Mdn=4.00). In the many sources conditions, we did not find a significant effect. We did not find other significant effects.

6.5.2 Effects of Number of Sources

As we have taken into account Hick's law (Hick 1952) to control for the amount of information, we can compare the mean values of the few and many sources conditions. We found that the participants need significantly more time per source to complete the tasks as the number of sources is tripled. The efficiency was highest for the no credibility rating/few sources condition (FN), where the average rank was 47.16 (M=214.26) and highest for the with credibility rating/few sources condition (FR) with an average rank of 50.29 (M=227.07). Our a priori assumption is confirmed. Moreover, the confidence drops, but not significantly, either when there is no credibility ratings (U=405.500, z=-.870, p(2-tailed)=.384), or with credibility ratings (U=393.500, z=-1.038, p(2-tailed)=.299), so we cannot support Hypothesis 1.

With the M-W U test, we found that the reliance on the content was significant when there was no credibility ratings (U=301.00, z=-2.495, p(2-tailed)=.013). Participants who had information from 4 sources had an average rank of 36.29 (Mdn=4.00) and participants who had 12 sources to assess had an average rank of 25.53 (Mdn=4.00), and hence relied less on the given content. We did not find more significant effects.

6.5.3 Effects Between Source and Content

Participants answered several statements about what their decisions were based on: the content (Q2) or the source (Q1). For the with credibility ratings conditions (FR and MR), an extra option was added: the visualization (Q3). The bar chart in Figure 6.3 shows the means from a 5-scale Likert value from 2 statements.

It shows that our participants agree more with statement 2 than statement 1 for all conditions. We check for statistical significant (2-tailed) correlation between these 2 variables for all 4 conditions with Spearman's rho. We found a large negative correlation for condition 1 (r=-.60, n=31, p<.0001), no correlation for condition 2 (r=-.01, n=30, p<.95), a significant medium negative correlation in condition 3 (r=-.49, n=31, p<.006), and a medium but not significant relationship in condition 4 (r=-.34, n=30, p<.06). The results show that there is a signifi-

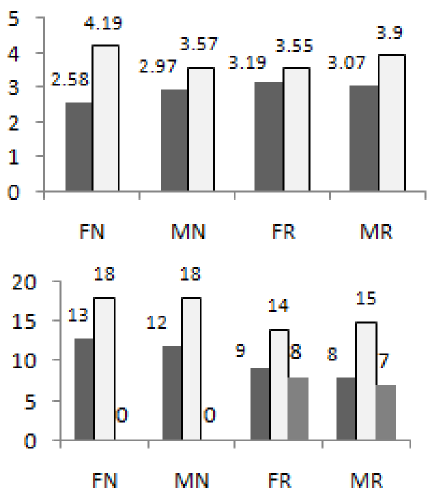


Figure 6.3: *Top:* Mean decision score per condition based on Q1 - Reliance on the source (black bar) or Q2 - Reliance on the Content (white bar) Score Min:1 Max:5
Bottom: The number of people that based decision on either: source (black bar), content (white bar) or visualization (grey bar). They are grouped by conditions.

cant strong relationship between source and answer for choosing information when people are confronted with the few sources conditions.

The bar chart in Figure 6.3 depicts what our participants choose for **Q3 Reliance on visualization**, which resulted in nominal data. Participants are asked to prioritize between the 2 factors that we focus on, and additionally the 3rd factor "the source credibility" is introduced in the *with source credibility* conditions. It shows that a majority of participants explicitly gave a higher priority to the *answers* in all conditions, which indicates that a large part of the participants preferred to use common sense, instead of blindly relying on the visualization of the source credibility. It also shows that participants were almost equally divided when choosing for either the source or the visualization of its credibility. However, we did not find this to be significant.

According to (Rieh and Danielson 2007), previous research has shown that credible sources are seen as likely to produce credible content, and credible contents are likely seen as to have originated from credible sources. The results of our study also show this clear tendency, where participants tend to select the highest ratings, however, this was also not significant.

6.5.4 Qualitative Feedback

In our studies, we also solicited user feedback on the experience and thoughts about the study they participated in. Making the source credibility ratings available significantly improved the confidence, however, some people do not rely blindly on it or understood it fully:

- *"The reliability meter seduces people to herd behavior."* [P26]
- *"For the first answer I trusted the Wikipedia and Amsterdam museum site, but since the answers for question 2 and further didn't make sense for these sites, I started looking only at the answers and not the sources."* [P31]
- *"I would explain the confidence bar more thoroughly at the start of the experiment. I understood it eventually, but it took some time."* [P72]

Some participants commented that the tasks given for both studies are difficult, especially the case with the many sources and no visualization condition. Our results show that this low confidence is significantly dependent on the cultural heritage knowledge of a participant, which is lacking for most of our novice participants.

- *"…average respondent gets their culture from NGC or Discovery Channel for no longer than average 10 minutes a week."* [P12]
- *"Whoa. I don't know anything about Japanese or Chinese art. I might just as well done a random selection..."* [P108]

6.6 Conclusions and Future Work

This study contributes to the scant research on interaction with applications that aggregate information from multiple sources. We investigated how displaying source credibility ratings influences people's behavior when accessing information. Our findings also suggest that performance is negatively influenced when users have to select information from many sources as compared with few sources, as people relied less on the given information, were less confident, and needed significantly more time as the number of sources increased. We also found that presenting credibility ratings do boast the confidence level for novice users. By making the source credibility ratings available, users can select information more confidently, even though people do not blindly rely on it.

In our study, it took participants more time to choose an answer in the conditions where the credibility ratings of the source was displayed. We expected the participants to lean heavily on the credibility indication to select (what seemed to them) the most reliable information, thus reducing search time. This is, however, not the case. The credibility indications may add to the user's cognitive load, or even that participants find the vizualisation a useful piece of information and choose to examine it longer.

Our research has prompted many follow up questions for future work. There is still a lot to be investigated on the topic of credibility ratings. For example, source credibility measures across different domains; In our experiment, we have anchored credibility rating to the 4 credibility measures for the Cultural Heritage domain (trustworthiness, completeness, goodwill and expertese). It would be interesting if there were other measures that are domain specific for example for the health, law, news, or financial domains.

Moreover, we question whether these credibility measures are the same for novice as well as expert users. We know from previous study (Amin et al. 2008) that experts have their own standards on which information sources that they trust and use.

Another question is how we can elicit credibility assessments of sources in a cheap but accurate way, given the plethora online sources in numerous domains. Examples of credibility scores assessments are: by user votes, by dedicated credibility assessors, by most visited, or by most linked/referred, etc. Should there be different assessment methods for different domains?

Furthermore, in our study, we manipulated the actual information to make them all equally plausible, in the real world low credible information may come from a high credible source, and vice versa. It will be interesting to investigate the choice behavior when a user is confronted with contradictions.

Finally, visualization of credibility ratings; There are different ways to visualize credibility, such as presenting the scores as numbers, bars, stars or other visualizations. Alternatively, it is also possible to visualize a stamp of approval for information sources that passed certain criteria. In our research, we only considered credibility visualization in a list result type presentation. Different types of result presentation, such as map view in `grokker.com`, may require other credibility visualization techniques. This raise the question whether the type of credibility visualization matters at all and if so, how does it influences users' search behavior. As online search systems act more as intelligent information aggregators, source credibility issues will be more prominent and will demand more attention from the Web community.

CHAPTER 7

Discussion and Conclusion

The work reported in the first part of this thesis provided insights into the information needs of both cultural heritage experts and mobile lay users when consulting multiple information sources. This allowed us to identify specific tasks that can be better supported by new information access tools. The studies carried out in the second part of this thesis allowed us to examine the effects of specific interface support for certain tasks. In this chapter we take a step back to reflect on the results obtained as well as the methods used. The chapter first addresses the research questions, then presents a discussion on the research methods, and finally we list suggestions for further research.

7.1 Research questions revisited

Research Question 1. What are the information seeking tasks for cultural heritage experts and mobile lay users?

We carried out two ethnographic studies in the domains of cultural heritage and location-based mobile search. In our study with cultural heritage experts, we conducted a semi-structured interview focused on professional information seeking tasks in online and offline sources (chapter 2). In our study with mobile lay users, we conducted a longitudinal digital diary study primarily focused on location based mobile search activities (chapter 3). Based on these, we discuss general trends in information seeking.

Cultural heritage experts and mobile lay users conduct both simple and complex information seeking tasks — Based on our studies, we conclude that cultural heritage experts and mobile lay users both conduct simple (e.g. fact finding and non-goal oriented tasks) and complex (e.g. information gathering) tasks. In our studies, we found that instances of information gathering tasks, such as comparison search, relation search, topic search, exploratory search and combinations of these, exist for both user groups (sections 2.4.3 and 3.3.1). However, the task distribution differs per user group. Cultural heritage experts tend to emphasize information gathering tasks. This is because cultural heritage experts' tasks tend to be more complex and heavily oriented towards analyzing information (section 2.4.3). The task distribution of mobile lay users is less focused (section 3.3.1). Many of the mobile lay users' tasks are simple, such as looking for business contact information. However, we also identified many complex information gathering tasks, such as comparing different businesses and products. Even though current mobile devices may not yet be suitable for complex information gathering tasks, we believe the frequency of these tasks will increase along with the increase of smartphone capabilities. Tasks usually currently done on a desktop will then also be performed on smartphones (section 3.4).

Cultural heritage experts and mobile lay users search across multiple sources — Based on our surveys of the information access tools used by cultural heritage experts and mobile lay users, we found that both types of users search in many information sources, sometimes simultaneously, to find their answers. Mobile lay users use different websites as information sources (section 3.4). Cultural heritage experts use online sources (reputable websites, collection management systems, search engines) as well as offline sources (research literature, catalogs and traditional archives) (section 2.4.2). When dealing with multiple sources, users have to manually aggregate and analyze scattered information to be able to synthesize their answer.

Based on our interviews, we conclude that the search tools of both cultural heritage experts and mobile lay users provide insufficient support for complex information seeking and multi-source search tasks. Comparison search is a good example where the two problems are combined in a single task. Comparison search applications typically support the user by aggregating and aligning similar pieces of information, and present these pieces in a way that makes it easy to compare results. In the cultural heritage domain, these laborious activities need to be done manually. Tasks become only more cumbersome when users need to compare results from different sources (section 5.3.3).

While our studies are only on two specific user groups and two domains, literature suggests that these trends and challenges are also applicable to other domains (section 2.4). In the following section we outline the requirements of information access tools addressing these challenges.

Research Question 2. What are the requirements to support information seeking across multiple sources?

We found that for most complex information seeking tasks across multiple sources, the difficulties primarily lie in inadequate support tools (sections 2.5.4, 3.5, 5.3.2). As a result, we identify two general requirements for information access tools: supporting simple and complex information seeking tasks and supporting search across multiple sources.

Support simple and complex information seeking tasks — Due to the different requirements of individual information seeking tasks, to fully support these tasks, different interfaces tailored to different tasks should be made available on top of the same data (2.5.4).

- Support fact finding and non-goal oriented tasks.
  Compared to the more complex search tasks discussed below, current tools tend to provide relatively good support for simple fact finding tasks. From our studies, however, we find there is still room for improvements. These improvements tend to be application and domain specific. In the cultural heritage domain, fact finding tasks can be better supported by systems that assist domain experts to do query expansion to help them build better queries, especially when dealing with unfamiliar museum collections (section 2.5). In the location-based mobile search domain, fact finding can be better supported with location and business-specific query refinement features (section 3.5). For non-goal oriented tasks in the cultural heritage domain, we found that technological solutions to support non-goal oriented tasks, such as RSS feeds, are already available even though not widely used. The main requirement here is to provide the most recent and relevant information while not letting the user feel overwhelmed by too much information, which turns out to be a difficult balance in practice (section 2.5.3). For the mobile domain, non-goal oriented tasks require better support in terms of recommendation based on people's social network and regular travel routes (section 3.6).

- Support information gathering tasks.
  The focus of this thesis is on complex information gathering tasks. Based on our survey of tools used, systems that support such tasks currently do not yet exist in the cultural heritage domain (section 2.5.4). More specifically, cultural heritage experts would appreciate tools that support tasks such

as comparison search, relation search, topic search, exploration search, and combinations of these (sections 2.4.3, 3.3.1 and 5.3.3). Generally, the challenge that users experience in carrying out information gathering tasks lies in finding higher-level relationships among individual facts that are distributed across and aggregated information from isolated search results (section 2.4.3). Therefore the application needs to provide links between the results from isolated searches and the aggregation support. More specific requirements, however, depend on the specific task. For example, in chapter 5 we investigated the requirements for comparison search tasks for cultural heritage experts. Here, the requirements are directly related to the experts' need to search, select and organize sets of artworks and to compare groups and individual artworks by finding similarities and differences among their properties and property values (section 5.4).

Support search across multiple sources— We found that there are different levels of requirements for applications that support information seeking across multiple sources.

- Provide aggregation support across multiple sources.
  The most basic problem in information seeking across multiple sources is that each source is a separate system that provides its own interface. Thus, to access information stored in different sources, users spend large amounts of time and effort on repetitive searching because they need to go to every source and repeat the same query or think about all possible alternative queries. This proves to be tedious and laborious work even for professionals (sections 2.5.4 and 5.3.2). Additionally, problems that occur while searching through a single source also occur while searching through multiple sources, but on an even larger scale. For example, in the cultural heritage domain, most problems with search within a single collection are related to the use of different name variants, different languages and different related terms. These problems only become worse when searching across multiple collections, where the information tends to be even more heterogeneous. Thus, one requirement for an aggregation support system is to provide reliable alignments between the different terms used by the different sources. By doing so, it should be possible to search on any alternative terms (section 5.4). On a related issue, to help users formulate their query, solutions that support users with finding the most appropriate terms should be provided (section 2.5.4 and 5.4). For example, in the location-based mobile search domain, one requirment would be that aggregation tools support the mashup of different types of informa-

tion, such as business information, product information, transportation and traffic information (section 3.6).

- Provide analysis support across multiple sources.
  In complex tasks, such as information gathering, it is not sufficient to aggregate and present individual search results. Analysis tools built on top of the data from multiple sources could provide added value. An example from the comparison search task in the cultural heritage domain is to provide support for users to analyze different properties and values of artworks belonging to different collections (section 5.4). Since different museums have their own conventions on how to annotate their collections, two requirements to provide analysis support are to find ways to align the information from the different collections accurately and to present them in the interface in a way that helps the expert carry out the comparison analysis effectively.

- Enable transparency of information sources.
  In a search aggregator, where many different information sources are combined, the user still needs to be able to assess how reliable the provided information is. For domain experts, source credibility is a crucial issue, and they typically only search in sources they trust (section 2.5.4). Generally these trusted sources are agreed upon within a community. For lay users, this consensus may not exist, and judging information source credibility can be difficult (section 6.6). For lay users to judge source credibility, systems are required to offer other means. For example, a system might provide recommendations from trusted people in the user's social network (section 3.6). One requirement for systems that use information from multiple sources is, therefore, to take credibility aspects into account, at the data, functional and interface level, to help users assess the reliability of the information presented.

Through our ethnography study, we have come across many inadequacies in the available search tools. To enable users to carry out their specific complex tasks better, these requirements should be addressed in future information access systems. In our third and final research question, we discuss alternative solutions that satisfy some of these requirements.

Research Question 3. How can we support information seeking tasks across multiple sources?

We carried out user studies directed towards finding solutions for advanced search tasks across multiple sources in the cultural heritage domain. We selected three areas to provide support: query formulation across multiple sources (chapters 4 and 5), analysis of aggregated search results from multiple sources (chapter 5) and transparency of information sources through credibility ratings (chapter 6).

Supporting query formulation across multiple sources— Within the cultural heritage domain, thesauri and other controlled vocabularies are commonly used to address problems related to synonyms, name variants, differences in languages or differences in jargon. Thesauri-based interfaces have, however, the reputation of being very hard to use, even for experts. We studied to what extent the ease of use of autocompletion-based interfaces could be combined with a set of aligned thesauri to support query formulation across multiple sources.

In chapter 5 we found that users can easily understand and use the tested interface, that it can save time when searching for many documents, and that by exploiting *synonym* relationships in the thesauri, it allows users to search and find documents that have been annotated with different terminology (section 5.6.3).

The key challenge in autocompletion usability is to manage the large number of potential suggestions. In chapter 4 we studied different strategies using the *hierarchical* and *type* relationships in the thesauri to organize autocompletion suggestions. We found that different thesauri require different strategies. The geographic thesaurus has symmetric levels and the grouping made by these levels were easy to understand. These grouping organizations were prefered by users over flat lists. They increase user performance by reducing typing errors and improve the speed of query formulation (section **??**). This was the not the case for the other thesaurus in the experiment. The lexical thesaurus has asymmetric levels. The groupings were difficult to understand. Thus, it is hard to make a good grouping and to predict which branch will contain the most appropriate suggestion. For this thesaurus, other means of organizing suggestions, such as alphabetically ordered lists or lists ranked on popularity, performed better (section 4.4).

In addition, we found that this feature requires sufficient data quality in terms of the annotations and the thesaurus alignments, a problem we discuss further in detail in section 7.1. Finally, we learned it is crucial to combine thesaurus-based autocompletion with free text search. When the user's query cannot be expressed by the predefined terms in the thesauri and the dataset included terms beyond those in the thesauri, free text search should also be available.

Supporting analysis of aggregated search results from multiple sources — Because different information gathering tasks have different requirements, we focussed on supporting a specific task, one frequently mentioned by the experts in chapter 2: comparison search. We designed a prototype application supporting this task and evaluated it using cultural heritage experts (chapter 5).

We developed the search module of the application by reusing the aligned thesauri and the autocompletion interface components of the previous study (chapter 4) and focussed on supporting the users with aggregation and (quantitative) analysis of the aggregated search results. The use of synonyms, name variants and other differences in terminology not only makes finding objects from multiple information sources harder, it also makes it hard to determine the aggregated properties of a group of search results, or to compare such groups. Inspired by this, we explored another use of thesauri alignments: how to support comparison search across aggregated information from multiple sources that are linked by thesauri. To support the comparison search task, we use different types of visualizations using the aggregated and partially linked information. We observed that users were generally positive about the experience with the novel tool and some of the visualizations were easy to use and allowed users to work effectively, in particular for complex comparison tasks (section 5.6.3). In general, the tool demonstrated that the use of thesauri can help to support a comparison task that is currently very time consuming and difficult to do manually. We found that the data quality of the object metadata and thesaurus alignments crucial in order to obtain reliable visualizations (section 5.7).

Support transparency of information sources through credibility ratings — We investigated the effect of visualizing information source credibility for lay users (chapter 6). We found that when lay users are confronted with information coming from multiple sources, they largely rely on judging the content presented (section 6.5) or rely on recommendations from people in their social networks (section 3.5). We collected different credibility ratings of different types of cultural heritage sources and showed these ratings to lay users. We found that visualization of the credibility ratings improved the confidence of lay users in the selected information (section 6.5). This indicates that providing source credibility cues, such as ratings, is useful to guide lay users in assessing which source to trust. The issue of credibility of an information source in an information access tool is highly context dependent 6.2. Thus, investigation of the appropriate mechanism to convey credibility ratings of multiple sources should be tailored to the application. For example, in chapter 6, we used users' scores in a progress bar to reflect cultural heritage institutions' perceived credibility ratings. Depending on the domain and application, other means to derive credibility ratings (e.g. peer review or by authority) or other types of credibility cues, might be more appropriate.

In research question 3, we discussed various interface features developed to support complex search tasks in heterogeneous data from multiple sources. In the next section, we will discuss the challenges future research will face with regards to supporting information seeking tasks across multiple sources.

7.2 Research method revisited

While carrying out our research, we faced a number of challenges when evaluating applications for multiple sources (chapters 4 and 5). Even though the interface and interaction are the primary focus of this research, we found that there are two important factors to be considered: 1) eliminating the effects of the varying quality of the underlying data and the search algorithm (sections 4.4 and 5.7), and 2) the user interface evaluation method. These factors have implications for our application evaluations (van Ossenbruggen et al. 2008).

Reducing the effects of the varying quality of the underlying data and search algorithm — The thesaurus-based applications, developed in chapters 4 and 5, provide meaningful interactions with complex data for specific tasks. If the data comes from a single source, a relational database will suffice. All data and relationships between different data can be adjusted to meet the application requirements. This is not the case, however, in applications using data from multiple sources where both the instance data and the data schemas are beyond the control of the application developer. Consequently, issues that normally do not occur in single source applications, such as information duplication, inconsistencies and trust in the information sources, become a concern that influences the perceived benefit of the application.

- Lack of high quality test collections.
  Ideally, the data combined from multiple sources has to go through a quality evaluation and control process before being used in an application. This measure is important because as long as the quality of the data is not sufficient, the information presented may not make sense to users. In our case, low quality data could negatively influence the user interface evaluation (section 5.7). One way to isolate the influence of data quality is to use test collections that has been carefully validated. Unfortunately, to date, there is little consensus in the community on a test collection for semantic web applications evaluation. For most domains it is hard to find ready made datasets that meet this high quality criterion. We work mostly with cultural heritage datasets. In this domain, annotation of cultural heritage datasets is done conscientiously by domain experts to ensure a high quality of information (Hildebrand et al. 2009). Thus, we believe this domain provides one of the best and realistic datasets. We find, however, that even in this domain there are cases where

the quality of the datasets and their alignment is only partially sufficient for a user evaluation (section 5.7).

- Varying quality of search algorithms.
  The quality of the search and inference software is another important issue to consider in our evaluation. Advanced search and inference software is needed to allow efficient search and navigation through data from multiple sources. Although this is true for all applications providing information access, efficient search algorithms across multiple sources are still new and are in their infancy. In our research, we are constantly confronted with algorithms and heuristics that work well with one particular dataset, but perform poorly on another: approaches that work well with one ontological modeling choice, do not necessarily work for another. Ensuring the middleware is of sufficient quality for a realistic user study is a non-trivial task.

Because we could not totally isolate interface aspects from the underlying data and search software, in our evaluation we have *preselected specific tasks*, through informed trial and error (section 4.4), that use parts of the data that we know work well. With this approach, we try to eliminate any negative influence and distraction caused by low quality data or poor search performance. This enables us to evaluate the interface issues almost independently from the potentially poor quality data and search algorithm. This means, however, we cannot evaluate users performing any comparison search task in realistic setting because it might require another part of the data that is poor in quality.

Evaluating user interfaces for multiple sources — Our user interface evaluation focuses on two areas: evaluating unified interfaces on distributed data from multiple sources; and novel interfaces for tasks that are complex and not supported by current tools.

- Evaluating unified interfaces for data from multiple sources.
  An application that accesses multiple sources will behave differently from those that run on a single dataset. Typically the former would respond slower but provide more information in terms of volume. We faced this challenge when we wanted to carry out interface evaluation of our new application that accesses multiple sources and compare it with a baseline application that typically accesses a single source. To eliminate the influence of system response and information load and have a fair evaluation, we decided to load fewer datasets for the new application to be the same as the current application. The trade off is that our users did not experience accessing

different types of collections with the new application. Consequently, we missed the opportunity to gain any insights in this area.

- Evaluating new interfaces for tasks that are not supported by current tools. Our main challenge here is to conduct an evaluation to examine the value of such applications. Since there are no similar tools that help users with the same task, it is hard for us to set a suitable baseline in our evaluation. Our comparison search interface is a good example of this (chapter 5). The interface is designed to help users to search in multiple sources, and is equipped with many features that are currently not available in current search tools (section 2.4.2, 5.3.2). During our evaluation, we faced the difficulty of testing advanced and complex information seeking tasks on our new tool because those tasks cannot be carried out with the baseline tool. As a compromise, we decided to "downgrade" our evaluation and test only a limited set of tasks that we know were possible to do with the baseline interface. By doing so, we cannot fully evaluate the usefulness of the rich novel features that our new tool has to offer.

7.3 Future research

We feel that research on supporting information seeking tasks across multiple sources is still at an early stage. Based on the experiences throughout this research, we offer directions for future work in the area of Semantic Web and Human Computer Interaction.

Semantic Web
As discussed in earlier section (section 7.2), the success of user evaluations on semantic web applications is not independent of the data quality and back end process. Therefore, we encourage future work in the following areas:

- **Data quality improvements** — Trust in the information is crucial for our domain expert users, and low quality data may negatively influence users' evaluation of it. Thus, high quality data, in single as well as multiple sources, is an important requirement for our expert users (section 2.5.4). However, this is a challenge for applications that include multiple information sources, because the overall quality of the data in an application is the result of converting and combining different single sources. In particular, it depends on the quality of the original data from each source, the quality of the data format conversion, and the precision of the alignments between the different sources. Both the format conversion and alignment processes are likely to reduce the resulting data quality. For example, the alignment stage is fundamentally an imprecise process, since concepts that are aligned may not be identical but still sufficiently close to motivate creating an alignment. Further research is needed *to identify the characteristics of the different data quality problems (introduced at these different stages of data processing) of information from multiple sources and to investigate methods to systematically mitigate them.*

- **Systematic evaluation methods** — The work described in this thesis is part of the semantic web research done within the MultimediaN E-Culture project. As explained in section 7.2, interface evaluation on top of semantic web technology is challenging. Therefore, we believe that the semantic web community would benefit from a common agreement on the different types of evaluation, much like the efforts of the information retrieval community by organizing TREC and TRECVid conferences. To date, there is little consensus in the semantic web community on a common evaluation framework[1]. This includes three evaluation aspects: the quality of data process (conversion and alignment), the semantic search algorithms and the semantic web user interfaces. We encourage future work for the whole semantic web community to agree on different types of evaluation methods for these aspects. In particular, to:

  - Defining guidelines to measure the quality of data conversion and integration similar to those that have already been developed for evaluating the quality of vocabulary alignments (to evaluate data quality).
  - Establish benchmark tests to provide a more meaningful comparison between the different semantic search algorithms (to evaluate search algorithms)
  - Provide test collections, either 'ideal' dataset having high quality or a realistic dataset having medium quality, that researchers can use in application development and providing evaluation guidelines for interfaces working on top of the test collections (to evaluate user interfaces).

  We believe that establishing common evaluation methods for these aspects will ultimately result in higher quality semantic web data, search algorithms and user interfaces.

[1] The community has made a first step in this direction by setting up workshops, such as "International Workshop on Evaluation of Ontology-based tools" (EON2009) or the"Evaluation for Entity Search Track" as part of the Semantic Search 2010 Workshop. `http://km.aifb.kit.edu/ws/semsearch10/`

Human Computer Interaction
Despite the fact that the SWUI (Semantic Web User Interaction) community has existed since 2006, research on user interfaces for semantic web application is still in its infancy. Most of the research done in the semantic web community tends to focus on the semantic web infrastructure and back-end process. To date, there is little multidisciplinary research involving both semantic web search and human computer interaction. As a follow up to our research, we indicate neccessary future work for the human computer interaction community in the context of (semantic web) search applications.

- **Information seeking tasks taxonomy validation** — We conducted two ethnography studies on users' information seeking tasks for two user groups (domain experts and lay users) and two different domains (cultural heritage and mobile location-based search). Consequently, our findings on the information seeking tasks taxonomy were primarily based on these user groups (section 2.4.3 and 3.3.1.3). *Further validation of this taxonomy through investigation in other domains is a natural follow up for this work.*

- **Support information seeking tasks across multiple sources** — In earlier parts of this thesis, we described a number of ways to support information seeking tasks across multiple sources (sections 2.7, 3.7, 4.5, 5.8, 6.6 and 7.1). However, we feel that one of the most important future directions in this area is on *investigating applications to support different types of information gathering tasks*. The reason being that these tasks are among the least supported by most information access tools. As initial work, in this thesis we investigated means to support comparison search task 5. Similar work is needed for other information gathering tasks. For example, to support relation search, more investigation is needed in the area of relation search algorithms, such as finding solutions to relate data from multiple sources and to present only meaningful relationships according to specific user tasks (van Ossenbruggen et al. 2008), and information presentation, such as investigations to find better visualizations to present different relationships in a rich way while maintaining ease-of-use.

- **Understanding the trade-offs between data quality and data coverage for applications with multiple sources** — Ideally, applications should always use high quality and trustworthy datasets. In reality, due to data processing and the characteristics of the original sources, the more information sources are combined together, the more likely that the average quality of the overall dataset decreases. This is often the case when dealing with real datasets from multiple sources having different qualities, such as linked data on the Web. It would be interesting for the SWUI community to investigate where such data can be still be valuable. Future work should investigate *what are the types of information seeking tasks and use cases where users are willing to tolerate lower data quality in return for larger data coverage across multiple sources?*

Once these specific tasks and use cases have been identified, follow-up work for the SWUI community would be on *how to design an interface and interaction for this type of information in a transparent way that enables users to make assessments, to adapt their search strategy and ultimately to find what they are looking for?*

The work in this thesis gives some signposts for the large amount of work that lies ahead in this research area. While we have identified a number of requirements that need to be satisfied when providing support for complex information seeking tasks, we have been able to investigate only three examples. Nevertheless, we hope that the examples that have been developed will provide inspiration to the semantic web community to understand the vital role that reliable data play in an end-user application. In parallel, we encourage the human computer interaction (HCI) community to embrace the huge challenges of developing new functionality in the midst of unstable technology. While the development of semantic web is predominantly an under-the-hood technology, we believe we have shown that this brings with it real HCI challenges, and some indications of how these may be addressed.

References

Agichtein, E., E. Brill, and S. Dumais (2006). Improving web search ranking by incorporating user behavior information. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 19–26. ACM.

Amin, A., L. Hardman, J. van Ossenbruggen, and A. van Nispen (2008). Understanding cultural heritage experts' information seeking tasks. In *JCDL '08: Proceedings of the Joint Conference of Digital Library*, New York, NY, USA, pp. 39–47. ACM Press.

Amin, A., M. Hildebrand, J. van Ossenbruggen, V. Evers, and L. Hardman (2009). Organizing Suggestions In Autocompletion Interfaces. In *ECIR'09: Proceedings of the Annual European Conference on Information Retrieval*, pp. 521–529.

Amin, A., M. Hildebrand, J. van Ossenbruggen, and L. Hardman (2010). Designing a thesaurus-based comparison search interface for linked cultural heritage sources. In *IUI'10: International Conference on Intelligent User Interfaces*, pp. 249–258. ACM Press.

Amin, A., S. Townsend, J. van Ossenbruggen, and L. Hardman (2009). Fancy a drink in canary wharf?": a user study on location-based mobile search. In *Interact '09: 12th IFIP TC13 Conference in Human-Computer Interaction*, Uppsala, Sweden, pp. 736–749. ACM Press.

Amin, A., J. Zhang, H. Cramer, L. Hardman, and V. Evers (2009). The effects of source credibility ratings in a cultural heritage information aggregator. In *WICOW '09: Proceedings of the 3rd workshop on Information credibility on the web*, New York, NY, USA, pp. 35–42. ACM.

Amin, A. K., M. Hildebrand, J. R. van Ossenbruggen, V. Evers, and L. Hardman (2008, December). Organizing Suggestions In Autocompletion Interfaces. Technical report, CWI.

Aula, A. and D. M. Russell (2009, March). Complex and exploratory web search. Information Seeking Support Systems Workshop.

Baeza-Yates, R., G. Dupret, and J. Velasco (2007). A study of mobile search queries in Japan. In *Query Log Analysis: Social and Technological Challenges, at WWW 2007*.

Baldonado, M. Q. W. (2000). A user-centered interface for information exploration in a heterogeneous digital library. *JASIS 51*(3), 297–310.

Bast, H., D. Majumdar, and I. Weber (2007). Efficient interactive query expansion with complete search. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, New York, NY, USA, pp. 857–860. ACM.

Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation 53*, 8—19.

Berlo, D. K., J. B. Lemert, and R. J. Mertz (1970). Dimensions for Evaluating the Acceptability of Message Sources. *Public Opin Q 33*(4), 563–576.

Brandt, J., N. Weiss, and S. R. Klemmer (2007). txt 4 l8r: lowering the burden for diary studies under mobile conditions. In *CHI '07: CHI '07 extended abstracts on Human factors in computing systems*, New York, NY, USA, pp. 2303–2308. ACM.

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum 36*(2), 3–10.

Brown, B. and M. Chalmers (2003). Tourism and mobile technology. In *ECSCW'03: Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, Norwell, MA, USA, pp. 335–354. Kluwer Academic Publishers.

Bruce, H., W. P. Jones, and S. T. Dumais (2004). Information behaviour that keeps found things found. *Inf. Res. 10*(1), 207.

Callahan, E. and J. Koenemann (2000). A comparative usability evaluation of user interfaces for online product catalogs. In *EC '00: Proceedings of the 2nd conference on Electronic commerce*, New York, NY, USA, pp. 197–206. ACM.

Case, D. O. (2006, December). *Looking for Information, Second Edition: A Survey of Research on Information Seeking, Needs, and Behavior (Library and Information Science)*. Academic Press.

Cassidy, W. P. (2007). Online news credibility: An examination of the perceptions of newspaper journalists. *Journal of Computer-Mediated Communication 12*(2), 478–498.

Chesney, T. (2006, November). An empirical examination of Wikipedia's credibility. *First Monday 11*(11), 1.

Choi, J. H., J. H. Watt, and M. Lynch (2006). Perceptions of news credibility about the war in iraq: Why war opponents perceived the internet as the most credible medium. *Journal of Computer-Mediated Communication 12*(1), 209–229.

Choo, C. W., B. Detlor, and D. Turnbull (1998). A behavioral model of information seeking on the web: preliminary results of a study of how managers and it specialist use the web. In *ASIS Annual Meeting Contributed Paper*.

Choo, C. W., B. Detlor, and D. Turnbull (2000a). *Web Work: Information Seeking and Knowledge Work on the World Wide Web*. Norwell, MA, USA: Kluwer Academic Publishers.

Choo, W. C., B. Detlor, and D. Turnbull (2000b). Information seeking on the web: An integrated model of browsing and searching. *First Monday 5*(2), 1.

Church, K., B. Smyth, P. Cotter, and K. Bradley (2007). Mobile information access: A study of emerging search behavior on the mobile internet. *ACM Trans. Web 1*(1), 4.

Consolvo, S. and M. Walker (2003). Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing 2*(2), 24–31.

Craigie, M., B. Loader, R. Burrows, and S. Muncer (2002, Jan). Reliability of health information on the internet: An examination of experts' ratings. *J Med Internet Res 4*(1), e2.

Cramer, H., V. Evers, S. Ramlal, M. Someren, L. Rutledge, N. Stash, L. Aroyo, and B. Wielinga (2008). The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction 18*(5), 455–496.

Cui, Y. and V. Roto (2008). How people use the web on mobile devices. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, New York, NY, USA, pp. 905–914. ACM.

Duff, W., B. Craig, and J. Cherry (2004). Historians' use of archival sources: Promises and pitfalls of the digital age. *Public Historian 26*(2), 7–22.

Eastin, M. S. (2001). Credibility assessments of online health information: The effects of source expertise and knowledge of content. *Journal of Computer-Mediated Communication 6*(4), 0–0.

Efthimiadis, E. N. (2000). Interactive query expansion: a user-based evaluation in a relevance feedback environment. *J. Am. Soc. Inf. Sci. 51*, 989–1003.

Endig, M., M. Hding, G. Saake, K.-U. Sattler, and E. Schallehn (2000). Federation services for heterogeneous digital libraries accessing cooperative and non-cooperative sources. In *Kyoto International Conference on Digital Libraries*, pp. 314–321.

Eysenbach, G. and C. Kohler (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *BMJ 324*(7337), 573–577.

Fisher, K. E., S. Erdelez, and L. E. Mckechnie (2005, August). *Theories of Information Behavior.* Asist Monograph. Information Today.

Flanagin, A. J. and M. J. Metzger (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media Society 9*(2), 319–342.

Fogg, B. J. and H. Tseng (1999). The elements of computer credibility. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 80–87. ACM Press.

Gonzales, J. I. B. (1994). A theory of organization. In *SIGDOC '94: Proceedings of the 12th annual international conference on Systems documentation*, New York, NY, USA, pp. 145–155. ACM.

Hendrickson, J. J. (1989). Performance, preference, and visual scan patterns on a menu-based system: implications for interface design. In *CHI '89: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 217–222. ACM.

Hick, W. E. (1952). On the rate of gain of information. *The Quarterly Journal of Experimental Psychology 4*(1), 11–26.

Hildebrand, M., J. van Ossenbruggen, and L. Hardman (2006). /facet: A browser for heterogeneous semantic web repositories. In *The Semantic Web - ISWC 2006*, Volume Volume 4273/2006 of *Lecture Notes in Computer Science*, pp. 272–285. Springer Berlin / Heidelberg.

Hildebrand, M., J. van Ossenbruggen, and L. Hardman (2009, Oct). Supporting subject matter annotation using heterogeneous thesauri: A user study in web data reuse. *International Journal of Human-Computer Studies 67*, 887–902.

Hildebrand, M., J. R. van Ossenbruggen, A. K. Amin, L. Aroyo, J. Wielemaker, and L. Hardman (2007). The Design Space Of A Configurable Autocompletion Component. Technical Report INS-E0708, CWI.

Hodgson, G. M. and S. R. Ruth (1985). The use of menus in the design of on-line sytems: a retrospective view. *SIGCHI Bull. 17*(1), 16–22.

Hong, T. (2006). The influence of structural and message features on web site credibility. *J. Am. Soc. Inf. Sci. Technol. 57*(1), 114–127.

Hovland, Carl I. and Weiss, Walter (1951). The influence of source credibility on communication effectiveness. *The Public Opinion Quarterly 15*(4), 635–650.

Hyvonen, E., E. Makela, T. Kauppinen, O. Alm, J. Kurki, T. Ruotsalo, K. Seppala, J. Takala, K. Puputti, H. Kuittinen, K. Viljanen, J. Tuominen, T. Palonen, M. Frosterus, R. Sinkkila, P. Paakkarinen, J. Laitio, and K. Nyberg (2009, March). Culturesampo – Finnish cultural heritage collections on the semantic web 2.0. In *Proceedings of the 1st International Symposium on Digital humanities for Japanese Arts and Cultures (DH-JAC-2009), Ritsumeikan, Kyoto, Japan.*

Ingwersen, P. and K. Järvelin (2005). *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Jakobsson, M. (1986). Autocompletion in full text transaction entry: a method for humanized input. In *CHI '86: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 327–332. ACM.

Jansen, B. J., D. L. Booth, and A. Spink (2007). Determining the user intent of web search engine queries. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, New York, NY, USA, pp. 1149–1150. ACM.

Joho, H., C. Coverson, M. Sanderson, and M. Beaulieu (2002). Hierarchical presentation of expansion terms. In *SAC '02: Proceedings of the 2002 ACM symposium on Applied computing*, New York, NY, USA, pp. 645–649. ACM.

Joho, H., M. Sanderson, and M. Beaulieu (2004). A study of user interaction with a concept-based interactive query expansion support tool. In *ECIR 2004*, pp. 42–56.

Kamvar, M. and S. Baluja (2006). A large scale study of wireless search behavior: Google mobile search. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, New York, NY, USA, pp. 701–709. ACM.

Kamvar, M. and S. Baluja (2007). The role of context in query input: using contextual signals to complete queries on mobile devices. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, New York, NY, USA, pp. 405–412. ACM.

Kellar, M., C. Watters, and K. M. Inkpen (2007). An exploration of web-based monitoring: implications for design. In *CHI '07: Proc. of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 377–386. ACM Press.

Kellar, M., C. Watters, and M. Shepherd (2005). A field study characterizing web-based information seeking tasks. Technical report, Dalhousie University Halifax.

Kellar, M., C. Watters, and M. Shepherd (2006). The impact of task on the usage of web browser navigation mechanisms. In *GI '06: Proc. of the 2006 conference on Graphics interface*, Toronto, Ont., Canada, Canada, pp. 235–242. Canadian Information Processing Society.

Kelly, D. and N. J. Belkin (2001). Reading time, scrolling and interaction: exploring implicit sources of user preferences for relevance feedback. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 408–409. ACM.

Kelly, D., A. Cushing, M. Dostert, X. Niu, and K. Gyllstrom (2010). Effects of popularity and quality on the usage of query suggestions during information search. In *CHI '10: Proceedings of the 28th international conference on Human factors in computing systems*, New York, NY, USA, pp. 45–54. ACM.

Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication & Society 4*(4), 381–403.

Kobsa, A. (2001). An empirical comparison of three commercial information visualization systems. In *INFOVIS '01: Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, Washington, DC, USA, pp. 123. IEEE Computer Society.

Koenemann, J. and N. J. Belkin (1996). A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 205–212. ACM.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the users perspective. *Journal of the American Society for Information Science 42*, 361–371.

Lee, E. S. and D. R. Raymond (1993). Menu-driven systems. In A. Kent and J. G. Williams (Eds.), *Encyclopedia of Microcomputers*, Volume 11, pp. 101–127. Marcel Dekker, New York.

Lee, J., H. S. Lee, and P. Wang (2004). An interactive visual interface for online product catalogs. *Electronic Commerce Research 4*(4), 335–358.

Lindberg, D. A. B. and B. L. Humphreys (1998). Medicine and Health on the Internet: The Good, the Bad, and the Ugly. *JAMA 280*(15), 1303–1304.

Lopes, R. and L. Carriço (2008). On the credibility of Wikipedia: an accessibility perspective. In *WICOW '08: Proceeding of the 2nd ACM workshop on Information credibility on the web*, New York, NY, USA, pp. 27–34. ACM.

Lynch, K. (1960). *The image of the city.* The M.I.T. Press, Cambridge, MA.

Manolescu, I., D. Florescu, and D. Kossmann (2001). Answering XML queries on heterogeneous data sources. In *VLDB '01: Proceedings of the 27th International Conference on Very Large Data Bases*, San Francisco, CA, USA, pp. 241–250. Morgan Kaufmann Publishers Inc.

Mark, G., U. Christensen, and M. Shafae (2001). A methodology using a microcamera for studying mobile. In *IT Usage and Person Mobility, CHI'01 Workshop on Mobile Communications: Understanding Users, Adoption & Design.*

Metzger, M. J., A. J. Flanagin, and L. Zwarun (2003). College student web use, perceptions of information credibility, and verification behavior. *Comput. Educ. 41*(3), 271–290.

Morrison, J. B., P. Pirolli, and S. K. Card (2001). A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *CHI '01 ext. abstracts on Human factors in computing systems*, New York, USA, pp. 163–164. ACM Press.

Murdock, V. and M. Lalmas (2008). Workshop on aggregated search. *SIGIR Forum 42*(2), 80–83.

O'Hara, K. and M. Perry (2001). Shopping anytime anywhere. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, New York, NY, USA, pp. 345–346. ACM.

Perugini, S., K. McDevitt, R. Richardson, M. A. Pérez-Qui R. Shen, N. Ramakrishnan, C. Williams, and E. A. Fox (2004). Enhancing usability in citidel:

multimodal, multilingual, and interactive visualization interfaces. In *JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 315–324. ACM.

Peterson, G., P. Aslani, and A. K. Williams (2003, Dec). How do consumers search for and appraise information on medicines on the internet? A qualitative study using focus groups. *J Med Internet Res 5*(4), e33.

Pirolli, P. L. T. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press.

Plaisant, C., J. Rose, B. Yu, L. Auvil, M. G. Kirschenbaum, M. N. Smith, T. Clement, and G. Lord (2006). Exploring erotics in emily dickinson's correspondence with text mining and visual interfaces. In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 141–150. ACM.

Pyrounakis, G., K. Saidis, M. Nikolaidou, and I. Lourdi (2004). Designing an integrated digital library framework to support multiple heterogeneous collections. In R. Heery and L. Lyon (Eds.), *ECDL*, Volume 3232 of *Lecture Notes in Computer Science*, pp. 26–37. Springer.

Radlinski, F. (2005). Query chains: Learning to rank from implicit feedback. In *KDD*, pp. 239–248. ACM Press.

Riegelsberger, J. and Y. Nakhimovsky (2008). Seeing the bigger picture: a multimethod field trial of google maps for mobile. In *CHI '08: CHI '08 extended abstracts on Human factors in computing systems*, New York, NY, USA, pp. 2221–2228. ACM.

Rieh, S. Y. and D. R. Danielson (2007). Credibility: A multidisciplinary framework. In *Annual Review of Information Science and Technology*, pp. 307–364.

Rosch, E. (1988). Principles of categorization. *Cognitive Science, a Persepective from Psychology and Artificial Intelligence 1*, 27–48.

Rose, D. E. and D. Levinson (2004). Understanding user goals in web search. In *WWW '04: Proc. of the 13th international conference on World Wide Web*, New York, NY, USA, pp. 13–19. ACM Press.

Rosner, D. K., J. M. Josling, A. Moed, and E. Oreglia (2007). Understanding target users of a digital reference library. In *JCDL '07: Proceedings of the 2007 conference on Digital libraries*, New York, NY, USA, pp. 495–495. ACM.

Sanderson, M. and Y. Han (2007). Search words and geography. In *GIR '07: Proceedings of the 4th ACM workshop on Geographical information retrieval*, New York, NY, USA, pp. 13–14. ACM.

Sanderson, M. and J. Kohler (2004). Analyzing geographic queries. In *Workshop on Geographic Information Retrieval SIGIR*.

Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen (2007). Collaborative filtering recommender systems. pp. 291–324.

Schreiber, G., A. Amin, L. Aroyo, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, B. Omelayenko, J. van Ossenbruggen, A. Tordai, J. Wielemaker, and B. J. Wielinga (2008). Semantic annotation and search of cultural-heritage collections: The multimedian e-culture demonstrator. *J. Web Sem. 6*(4), 243–249.

Schreiber, G., A. Amin, M. van Assem, V. de Boer, L. Hardman, M. Hildebrand, L. Hollink, Z. Huang, J. van Kersen, M. de Niet, B. Omelayenko, J. van Ossenbruggen, R. Siebes, J. Taekema, J. Wielemaker, and B. Wielinga (2006). Multimedian e-culture demonstrator. In I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo (Eds.), *The Semantic Web - ISWC 2006*, Volume 4273 of *LNCS*, pp. 951–958. Springer.

Sellen, A. J., R. Murphy, and K. L. Shaw (2002). How knowledge workers use the web. In *CHI '02: Proc. of the SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 227–234. ACM Press.

Shen, R., N. S. Vemuri, W. Fan, R. da S. Torres, and E. A. Fox (2006). Exploring digital libraries: integrating browsing, searching, and visualization. In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 1–10. ACM.

Si, L. and J. Callan (2005). Modeling search engine effectiveness for federated search. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 83–90. ACM.

Sohn, T., K. A. Li, W. G. Griswold, and J. D. Hollan (2008). A diary study of mobile information needs. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, New York, NY, USA, pp. 433–442. ACM.

Spenke, M., C. Beilken, and T. Berlage (1996). Focus: the interactive table for product comparison and selection. In *UIST '96: Proceedings of the 9th annual ACM symposium on User interface software and technology*, New York, NY, USA, pp. 41–50. ACM.

Spink, A., B. J. Jansen, D. Wolfram, and T. Saracevic (2002). From e-sex to e-commerce: Web search changes. *Computer 35*(3), 107–109.

Steiger, P. and M. Stolze (1997). Effective product selection in electronic catalogs. In *CHI '97: CHI '97 extended abstracts on Human factors in computing systems*, New York, NY, USA, pp. 291–292. ACM.

Sundar, S. S. and C. Nass (2001). Conceptualizing sources in online news. *Journal of Communication 51*(1), 52–72.

Teevan, J., E. Adar, R. Jones, and M. A. S. Potts (2007). Information re-retrieval: repeat queries in yahoo's logs. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 151–158. ACM.

Tenev, T. and R. Rao (1997). Managing multiple focal levels in table lens. In *INFOVIS '97: Proceedings of the 1997 IEEE Symposium on Information Visualization (InfoVis '97)*, Washington, DC, USA, pp. 59. IEEE Computer Society.

Theus, M. and S. Urbanek (2009). *Interactive Graphics for Data Analysis*. Computer Science and Data Analysis.

Tibbo, H. R. (2002). Primarily history: historians and the search for primary source materials. In *JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 1–10. ACM.

Tordai, A., B. Omelayenko, and G. Schreiber (2007). Semantic excavation of the city of books. In *SAAKM*.

Tordai, A., J. van Ossenbruggen, and G. Schreiber (2009). Combining vocabulary alignment techniques. In *K-CAP '09: Proceedings of the fifth international conference on Knowledge capture*, New York, NY, USA, pp. 25–32. ACM.

Torres, R., S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl (2004). Enhancing digital libraries with techlens+. In *JCDL '04: Proc. of the 4th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 228–236. ACM Press.

Tseng, S. and B. J. Fogg (1999). Credibility and computing technology. *Commun. ACM 42*(5), 39–44.

Tsfati, Y. and J. N. Cappella (2005). Why do people watch news they do not trust? The need for cognition as a moderator in the association between news media skepticism and exposure. *Media Psychology 7*(3), 251–271.

van Assem, M., V. Malaisé, A. Miles, and G. Schreiber (2006). A method to convert thesauri to skos. *The Semantic Web: Research and Applications 4011*, 95–109.

van den Bosch, A., M. van Erp, and C. Sporleder (2009). Making a clean sweep of cultural heritage. *IEEE Intelligent Systems 24*(2), 54–63.

van der Graaf, M. (2006, Dec.). Behoefteonderzoek gebruikers collectiewijzer. Available at: http://project.collectiewijzer.nl/ images/stories/pdf/ Behoefteonderzoek_gebruikers_Collectiewijzer.pdf, last visited Dec 2007.

van Ossenbruggen, J., A. Amin, and M. Hildebrand (2008). Why evaluating semantic web applications is difficult. In *Semantic Web User Interaction Workshop, CHI'08*, pp. 1–4.

Walther, B. J., Z. Wang, and T. Loh (2004, Sep). The effect of top-level domains and advertisements on health web site credibility. *J Med Internet Res 6*(3), 24.

Warwick, C., J. Rimmer, A. Blandford, and G. Buchanan (2005). User centred interactive search in the humanities. In *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, New York, NY, USA, pp. 400–400. ACM.

White, R. W., M. Bilenko, and S. Cucerzan (2007). Studying the use of popular destinations to enhance web search interaction. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 159–166. ACM.

White, R. W. and G. Marchionini (2006). A study of real-time query expansion effectiveness. In *SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 715–716. ACM.

White, R. W. and G. Marchionini (2007). Examining the effectiveness of real-time query expansion. *Inf. Process. Manage. 43*(3), 685–704.

White, R. W. and D. Morris (2007). Investigating the querying and browsing behavior of advanced search engine users. In *SIGIR '07: Proc. of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, pp. 255–262. ACM Press.

Wielemaker, J., M. Hildebrand, J. Ossenbruggen, and G. Schreiber (2008). Thesaurus-based search in large heterogeneous collections. In *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*, Berlin, Heidelberg, pp. 695–708. Springer-Verlag.

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

Wilson, T. D. (2000). Human information behavior. *Information Science Research 3*, 49–55.

Yee, K. P., K. Swearingen, K. Li, and M. Hearst (2003). Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, Ft. Lauderdale, Florida, USA, pp. 401–408. ACM Press.

Yi, J., F. Maghoul, and J. Pedersen (2008). Deciphering mobile search patterns: a study of yahoo! mobile search queries. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, New York, NY, USA, pp. 257–266. ACM.

Young, F. W., P. M. Valero-Mora, and M. Friendly (2006). *Visual Statistics: Seeing Data with Dynamic Interactive Graphics*. Wiley Series.

Zhang, J. (2007). Finding cultural heritage information from multiple and diverse sources in semantic web applications: The effect of visualizing source credibility. Master's thesis, University of Amsterdam.

Zhang, J., A. Amin, H. Cramer, L. Hardman, and V. Evers (2009). Improvings user's confidence in cultural heritage aggregated results. In *SIGIR '09: Proceedings of the 32th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA. ACM.

Zhang, J., K. N. Fachry, and J. Kamps (2008). Access to archival finding aids: Context matters. In *ECDL '08: Proceedings of the 12th European conference on Research and Advanced Technology for Digital Libraries*, Berlin, Heidelberg, pp. 455–457. Springer-Verlag.

Summary

Driven by the explosion of information sources available on the Web, more people than ever rely on accessing online search applications to find answers buried in large numbers of documents. Because of this, research on information seeking behavior and how people use search applications is increasingly important. We now understand that there are many variants of information seeking tasks (Kellar et al. 2005), for example a simple search task: “I need the phone number of that restaurant.”, or a complex and vague one: “I need to find a good theme for the next museum exhibition.” These different tasks can be characterised by different aspects, such as: user types, e.g. experts vs. lay people and different domains, e.g. cultural heritage vs. e-commerce. Considering the plethora of information sources available online, search engines plays an important role in helping users find information from these sources.

With the ability to search across multiple sources, however, new challenges and problems arise, for example, how should information be presented in a way that users can understand the difference between different sources, or how can users navigate through multiple (unfamiliar) information sources to find their answers. End-user search across multiple sources is acknowledged to be a nontrivial problem by the human computer interaction community, e.g. (Aula and Russell 2009; Baldonado 2000), and remains a topic that has yet to be fully understood.

In this thesis, we investigate information seeking tasks across multiple sources and how they are influenced by different user types and different domains. The discussion is divided into two parts. The first part of the thesis presents two studies that contribute to understanding users’ information seeking tasks. Two different user types, domain experts and lay users, and two domains, cultural heritage and location based mobile search domain were used as case studies. The second part of the thesis discusses design requirements and examples of novel interfaces to support specific aspects of information seeking tasks across multiple sources.

The contributions of the research described in this thesis include:

- an in-depth analysis of information seeking tasks across multiple information sources for experts and lay users;

- functional and user interface requirements for future information access tools;
- reflections on the user interface evaluation methods used throughout the research
- challenges related to the implementation and evaluation of search applications for multiple sources using Semantic Web techologies.

The work in this thesis provides signposts for the large amount of work that lies ahead in this multidisciplinary research area. We identified a number of requirements that need to be satisfied when providing support for complex information seeking tasks and we have been able to investigate a few example applications. Nevertheless, we hope that the examples that have been developed will provide inspiration to the Semantic Web community as well as the Human Computer Interaction community. Through this thesis, we encourage both communities to embrace the huge challenges of developing user-centric applications for a fast growing technology.

Sammenvatting

Door de explosie van informatie op het web gebruiken meer mensen dan ooit online zoekapplicaties om antwoorden verborgen in stapels documenten te vinden. Vanwege deze ontwikkeling wordt het onderzoek naar zoekgedrag en naar hoe mensen gebruik maken van zoekapplicaties steeds belangrijker.

Uit onderzoek is gebleken dat er veel verschillende soorten van zoektaken zijn (Kellar et al. 2005), bijvoorbeeld een nauwkeurige zoekopdracht: "Ik heb het telefoonnummer van dat restaurant nodig." of een complexe en vage taak: "Ik moet een goed thema voor de volgende tentoonstelling van het museum vinden". Deze verschillende taken kunnen beïnvloed worden door andere aspecten zoals gebruikers-typen, bijvoorbeeld: een deskundige gebruiker of amateur; of domeinen, bijvoorbeeld: cultureel erfgoed of e-commerce. Door de overvloed aan beschikbare online informatie spelen zoekmachines een belangrijke rol om gebruikers de juiste informatie te helpen vinden.

Door de mogelijkheid om te zoeken in meerdere bronnen tegelijk, komen er nieuwe uitdagingen en problemen te voorschijn, bijvoorbeeld de vraag hoe de informatie moet worden gepresenteerd om een gebruiker goed inzicht in de verschillende herkomst te bieden, of de vraag hoe gebruikers kunnen navigeren in meerdere (onbekende) informatiebronnen om hun antwoorden te vinden. Zoeken in meerdere bronnen wordt erkend als een moelijk probleem door de Human Computer Interaction gemeenschap (Aula en Russell 2009; Baldonado 2000), een probleem dat we nog niet volledig hebben begrepen. In dit proefschrift worden zoektaken in meerdere bronnen onderzocht, alsmede de invloed van verschillende gebruikers-typen en domeinen.

Dit proefschrift bestaat uit twee delen. Het eerste deel van het proefschrift biedt inzicht in de zoektaken van verschillende gebruikers en domeinen. Als case studies werden twee gebruikers-typen gekozen, namelijk domein-deskundigen en onervaren gebruikers, en twee domeinen: cultureel erfgoed en mobiel zoeken. Het tweede deel van het proefschrift bespreekt het ontwerpproces en evaluatie van nieuwe, innovatieve interfaces om specifieke zoektaken te ondersteunen. De bijdrage van dit onderzoek is onder andere:

- Een diepe analyse van zoektaken in meerdere informatiebronnen voor deskundigen en onervaren gebruikers.
- Een lijst met functionele eisen voor toekomstige zoekmachines die zoektaken in meerdere bronnen uitvoeren.
- Refelecties over de interface-evaluatie-methoden die in dit onderzoek worden gebruikt, en een identificatie van de belangrijkste daarbij behorende uitdagingen.

Het werk in dit proefschrift biedt een aantal wegwijzers voor toekomstig onderzoek in dit multidisciplinaire veld. We hebben eisen geïdentificeerd om complexe zoektaken in meerdere bronnen te ondersteunen. Bovendien hebben we een aantal applicaties als voorbeeld ontwikkeld, en hopen dat deze voorbeeld-applicaties een inspiratie zullen zijn voor zowel de Semantic Web gemeenschap als de Human Computer Interaction gemeenschap.

SIKS Dissertation Series

1998

1 Johan van den Akker (CWI), DEGAS - An Active, Temporal Database of Autonomous Objects
2 Floris Wiesman (UM), Information Retrieval by Graphically Browsing Meta-Information
3 Ans Steuten (TUD), A Contribution to the, Linguistic Analysis of Business Conversations within the Language/Action Perspective
4 Ans Steuten (TUD), A Contribution to the Linguistic Analysis of Business Conversations within the Language/Action Perspective
5 Dennis Breuker (UM), Memory versus Search in Games
6 E.W.Oskamp (RUL), Computerondersteuning bij Straftoemeting

1999

1 Mark Sloof (VU), Physiology of Quality Change Modelling; Automated modelling of Quality Change of Agricultural Products
2 Rob Potharst (EUR), Classification using decision trees and neural nets
3 Don Beal (UM), The Nature of Minimax Search
4 Jacques Penders (UM), The practical Art of Moving Physical Objects
5 Aldo de Moor (KUB), Empowering Communities: A Method for the Legitimate User-Driven Specification of Network Information Systems
6 Niek J.E. Wijngaards (VU), Re-design of compositional systems
7 David Spelt (UT), Verification support for object database design
8 Jacques H.J. Lenting (UM), Informed Gambling: Conception and Analysis of a Multi-Agent Mechanism for Discrete Reallocation

2000

1 Frank Niessink (VU), Perspectives on Improving Software Maintenance
2 Koen Holtman (TUE), Prototyping of CMS Storage Management
3 Carolien M.T. Metselaar (UVA), Sociaal-organisatorische gevolgen van kennistechnologie; een procesbenadering en actorperspectief.
4 Geert de Haan (VU), ETAG, A Formal Model of Competence Knowledge for User Interface Design
5 Ruud van der Pol (UM), Knowledge-based Query Formulation in Information Retrieval.
6 Rogier van Eijk (UU), Programming Languages for Agent Communication
7 Niels Peek (UU), Decision-theoretic Planning of Clinical Patient Management
8 Veerle Coup (EUR), Sensitivity Analyis of Decision-Theoretic Networks
9 Florian Waas (CWI), Principles of Probabilistic Query Optimization
10 Niels Nes (CWI), Image Database Management System Design Considerations, Algorithms and Architecture
11 Jonas Karlsson (CWI), Scalable Distributed Data Structures for Database Management

2001

1 Silja Renooij (UU), Qualitative Approaches to Quantifying Probabilistic Networks
2 Koen Hindriks (UU), Agent Programming Languages: Programming with Mental Models
3 Maarten van Someren (UvA), Learning as problem solving
4 Evgueni Smirnov (UM), Conjunctive and Disjunctive Version Spaces with Instance-Based Boundary Sets
5 Jacco van Ossenbruggen (VU), Processing Structured Hypermedia: A Matter of Style
6 Martijn van Welie (VU), Task-based User Interface Design
7 Bastiaan Schonhage (VU), Diva: Architectural Perspectives on Information Visualization
8 Pascal van Eck (VU), A Compositional Semantic Structure for Multi-Agent Systems Dynamics.
9 Pieter Jan 't Hoen (RUL), Towards Distributed Development of Large Object-Oriented Models, Views of Packages as Classes
10 Maarten Sierhuis (UvA), Modeling and Simulating Work Practice BRAHMS: a multiagent modeling and simulation language for work practice analysis and design
11 Tom M. van Engers (VUA), Knowledge Management: The Role of Mental Models

in Business Systems Design

2002

1 Nico Lassing (VU), Architecture-Level Modifiability Analysis
2 Roelof van Zwol (UT), Modelling and searching web-based document collections
3 Henk Ernst Blok (UT), Database Optimization Aspects for Information Retrieval
4 Juan Roberto Castelo Valdueza (UU), The Discrete Acyclic Digraph Markov Model in Data Mining
5 Radu Serban (VU), The Private Cyberspace Modeling Electronic Environments inhabited by Privacy-concerned Agents
6 Laurens Mommers (UL), Applied legal epistemology; Building a knowledge-based ontology of the legal domain
7 Peter Boncz (CWI), Monet: A Next-Generation DBMS Kernel For Query-Intensive Applications
8 Jaap Gordijn (VU), Value Based Requirements Engineering: Exploring Innovative E-Commerce Ideas
9 Willem-Jan van den Heuvel(KUB), Integrating Modern Business Applications with Objectified Legacy Systems
10 Brian Sheppard (UM), Towards Perfect Play of Scrabble
11 Wouter C.A. Wijngaards (VU), Agent Based Modelling of Dynamics: Biological and Organisational Applications
12 Albrecht Schmidt (Uva), Processing XML in Database Systems
13 Hongjing Wu (TUE), A Reference Architecture for Adaptive Hypermedia Applications
14 Wieke de Vries (UU), Agent Interaction: Abstract Approaches to Modelling, Programming and Verifying Multi-Agent Systems
15 Rik Eshuis (UT), Semantics and Verification of UML Activity Diagrams for Workflow Modelling
16 Pieter van Langen (VU), The Anatomy of Design: Foundations, Models and Applications
17 Stefan Manegold (UVA), Understanding, Modeling, and Improving Main-Memory Database Performance

2003

1 Heiner Stuckenschmidt (VU), Ontology-Based Information Sharing in Weakly Structured Environments
2 Jan Broersen (VU), Modal Action Logics for Reasoning About Reactive Systems
3 Martijn Schuemie (TUD), Human-Computer Interaction and Presence in Virtual Reality Exposure Therapy
4 Milan Petkovic (UT), Content-Based Video Retrieval Supported by Database Technology
5 Jos Lehmann (UVA), Causation in Artificial Intelligence and Law - A modelling approach
6 Boris van Schooten (UT), Development and specification of virtual environments
7 Machiel Jansen (UvA), Formal Explorations of Knowledge Intensive Tasks
8 Yongping Ran (UM), Repair Based Scheduling
9 Rens Kortmann (UM), The resolution of visually guided behaviour
10 Andreas Lincke (UvT), Electronic Business Negotiation: Some experimental studies on the interaction between medium, innovation context and culture
11 Simon Keizer (UT), Reasoning under Uncertainty in Natural Language Dialogue using Bayesian Networks
12 Roeland Ordelman (UT), Dutch speech recognition in multimedia information retrieval
13 Jeroen Donkers (UM), Nosce Hostem - Searching with Opponent Models
14 Stijn Hoppenbrouwers (KUN), Freezing Language: Conceptualisation Processes across ICT-Supported Organisations
15 Mathijs de Weerdt (TUD), Plan Merging in Multi-Agent Systems
16 Menzo Windhouwer (CWI), Feature Grammar Systems - Incremental Maintenance of Indexes to Digital Media Warehouses
17 David Jansen (UT), Extensions of Statecharts with Probability, Time, and Stochastic Timing
18 Levente Kocsis (UM), Learning Search Decisions

2004

1 Virginia Dignum (UU), A Model for Organizational Interaction: Based on Agents, Founded in Logic
2 Lai Xu (UvT), Monitoring Multi-party Contracts for E-business
3 Perry Groot (VU), A Theoretical and Empirical Analysis of Approximation in Symbolic Problem Solving
4 Chris van Aart (UVA), Organizational Principles for Multi-Agent Architectures
5 Viara Popova (EUR), Knowledge discovery and monotonicity

6 Bart-Jan Hommes (TUD), The Evaluation of Business Process Modeling Techniques
7 Elise Boltjes (UM), Voorbeeldig onderwijs; voorbeeldgestuurd onderwijs, een opstap naar abstract denken, vooral voor meisjes
8 Joop Verbeek(UM), Politie en de Nieuwe Internationale Informatiemarkt, Grensregionale politiële gegevensuitwisseling en digitale expertise
9 Martin Caminada (VU), For the Sake of the Argument; explorations into argument-based reasoning
10 Suzanne Kabel (UVA), Knowledge-rich indexing of learning-objects
11 Michel Klein (VU), Change Management for Distributed Ontologies
12 The Duy Bui (UT), Creating emotions and facial expressions for embodied agents
13 Wojciech Jamroga (UT), Using Multiple Models of Reality: On Agents who Know how to Play
14 Paul Harrenstein (UU), Logic in Conflict. Logical Explorations in Strategic Equilibrium
15 Arno Knobbe (UU), Multi-Relational Data Mining
16 Federico Divina (VU), Hybrid Genetic Relational Search for Inductive Learning
17 Mark Winands (UM), Informed Search in Complex Games
18 Vania Bessa Machado (UvA), Supporting the Construction of Qualitative Knowledge Models
19 Thijs Westerveld (UT), Using generative probabilistic models for multimedia retrieval
20 Madelon Evers (Nyenrode), Learning from Design: facilitating multidisciplinary design teams

2005

1 Floor Verdenius (UVA), Methodological Aspects of Designing Induction-Based Applications
2 Erik van der Werf (UM), AI techniques for the game of Go
3 Franc Grootjen (RUN), A Pragmatic Approach to the Conceptualisation of Language
4 Nirvana Meratnia (UT), Towards Database Support for Moving Object data
5 Gabriel Infante-Lopez (UVA), Two-Level Probabilistic Grammars for Natural Language Parsing
6 Pieter Spronck (UM), Adaptive Game AI
7 Flavius Frasincar (TUE), Hypermedia Presentation Generation for Semantic Web Information Systems
8 Richard Vdovjak (TUE), A Model-driven Approach for Building Distributed Ontology-based Web Applications
9 Jeen Broekstra (VU), Storage, Querying and Inferencing for Semantic Web Languages
10 Anders Bouwer (UVA), Explaining Behaviour: Using Qualitative Simulation in Interactive Learning Environments
11 Elth Ogston (VU), Agent Based Matchmaking and Clustering - A Decentralized Approach to Search
12 Csaba Boer (EUR), Distributed Simulation in Industry
13 Fred Hamburg (UL), Een Computermodel voor het Ondersteunen van Euthanasiebeslissingen
14 Borys Omelayenko (VU), Web-Service configuration on the Semantic Web; Exploring how semantics meets pragmatics
15 Tibor Bosse (VU), Analysis of the Dynamics of Cognitive Processes
16 Joris Graaumans (UU), Usability of XML Query Languages
17 Boris Shishkov (TUD), Software Specification Based on Re-usable Business Components
18 Danielle Sent (UU), Test-selection strategies for probabilistic networks
19 Michel van Dartel (UM), Situated Representation
20 Cristina Coteanu (UL), Cyber Consumer Law, State of the Art and Perspectives
21 Wijnand Derks (UT), Improving Concurrency and Recovery in Database Systems by Exploiting Application Semantics

2006

1 Samuil Angelov (TUE), Foundations of B2B Electronic Contracting
2 Cristina Chisalita (VU), Contextual issues in the design and use of information technology in organizations
3 Noor Christoph (UVA), The role of metacognitive skills in learning to solve problems
4 Marta Sabou (VU), Building Web Service Ontologies
5 Cees Pierik (UU), Validation Techniques for Object-Oriented Proof Outlines
6 Ziv Baida (VU), Software-aided Service Bundling - Intelligent Methods & Tools for Graphical Service Modeling
7 Marko Smiljanic (UT), XML schema matching – balancing efficiency and effectiveness by means of clustering
8 Eelco Herder (UT), Forward, Back and

Home Again - Analyzing User Behavior on the Web
9 Mohamed Wahdan (UM), Automatic Formulation of the Auditor's Opinion
10 Ronny Siebes (VU), Semantic Routing in Peer-to-Peer Systems
11 Joeri van Ruth (UT), Flattening Queries over Nested Data Types
12 Bert Bongers (VU), Interactivation - Towards an e-cology of people, our technological environment, and the arts
13 Henk-Jan Lebbink (UU), Dialogue and Decision Games for Information Exchanging Agents
14 Johan Hoorn (VU), Software Requirements: Update, Upgrade, Redesign - towards a Theory of Requirements Change
15 Rainer Malik (UU), CONAN: Text Mining in the Biomedical Domain
16 Carsten Riggelsen (UU), Approximation Methods for Efficient Learning of Bayesian Networks
17 Stacey Nagata (UU), User Assistance for Multitasking with Interruptions on a Mobile Device
18 Valentin Zhizhkun (UVA), Graph transformation for Natural Language Processing
19 Birna van Riemsdijk (UU), Cognitive Agent Programming: A Semantic Approach
20 Marina Velikova (UvT), Monotone models for prediction in data mining
21 Bas van Gils (RUN), Aptness on the Web
22 Paul de Vrieze (RUN), Fundaments of Adaptive Personalisation
23 Ion Juvina (UU), Development of Cognitive Model for Navigating on the Web
24 Laura Hollink (VU), Semantic Annotation for Retrieval of Visual Resources
25 Madalina Drugan (UU), Conditional log-likelihood MDL and Evolutionary MCMC
26 Vojkan Mihajlovic (UT), Score Region Algebra: A Flexible Framework for Structured Information Retrieval
27 Stefano Bocconi (CWI), Vox Populi: generating video documentaries from semantically annotated media repositories
28 Borkur Sigurbjornsson (UVA), Focused Information Access using XML Element Retrieval

2007

1 Kees Leune (UvT), Access Control and Service-Oriented Architectures
2 Wouter Teepe (RUG), Reconciling Information Exchange and Confidentiality: A Formal Approach
3 Peter Mika (VU), Social Networks and the Semantic Web
4 Jurriaan van Diggelen (UU), Achieving Semantic Interoperability in Multi-agent Systems: a dialogue-based approach
5 Bart Schermer (UL), Software Agents, Surveillance, and the Right to Privacy: a Legislative Framework for Agent-enabled Surveillance
6 Gilad Mishne (UVA), Applied Text Analytics for Blogs
7 Natasa Jovanovic' (UT), To Whom It May Concern - Addressee Identification in Face-to-Face Meetings
8 Mark Hoogendoorn (VU), Modeling of Change in Multi-Agent Organizations
9 David Mobach (VU), Agent-Based Mediated Service Negotiation
10 Huib Aldewereld (UU), Autonomy vs. Conformity: an Institutional Perspective on Norms and Protocols
11 Natalia Stash (TUE), Incorporating Cognitive/Learning Styles in a General-Purpose Adaptive Hypermedia System
12 Marcel van Gerven (RUN), Bayesian Networks for Clinical Decision Support: A Rational Approach to Dynamic Decision-Making under Uncertainty
13 Rutger Rienks (UT), Meetings in Smart Environments; Implications of Progressing Technology
14 Niek Bergboer (UM), Context-Based Image Analysis
15 Joyca Lacroix (UM), NIM: a Situated Computational Memory Model
16 Davide Grossi (UU), Designing Invisible Handcuffs. Formal investigations in Institutions and Organizations for Multi-agent Systems
17 Theodore Charitos (UU), Reasoning with Dynamic Networks in Practice
18 Bart Orriens (UvT), On the development an management of adaptive business collaborations
19 David Levy (UM), Intimate relationships with artificial partners
20 Slinger Jansen (UU), Customer Configuration Updating in a Software Supply Network
21 Karianne Vermaas (UU), Fast diffusion and broadening use: A research on residential adoption and usage of broadband internet in the Netherlands between 2001 and 2005
22 Zlatko Zlatev (UT), Goal-oriented design of value and process models from patterns
23 Peter Barna (TUE), Specification of Application Logic in Web Information Systems
24 Georgina Ramrez Camps (CWI), Structural Features in XML Retrieval

25 Joost Schalken (VU), Empirical Investigations in Software Process Improvement

2008

1 Katalin Boer-Sorbn (EUR), Agent-Based Simulation of Financial Markets: A modular,continuous-time approach
2 Alexei Sharpanskykh (VU), On Computer-Aided Methods for Modeling and Analysis of Organizations
3 Vera Hollink (UVA), Optimizing hierarchical menus: a usage-based approach
4 Ander de Keijzer (UT), Management of Uncertain Data - towards unattended integration
5 Bela Mutschler (UT), Modeling and simulating causal dependencies on process-aware information systems from a cost perspective
6 Arjen Hommersom (RUN), On the Application of Formal Methods to Clinical Guidelines, an Artificial Intelligence Perspective
7 Peter van Rosmalen (OU), Supporting the tutor in the design and support of adaptive e-learning
8 Janneke Bolt (UU), Bayesian Networks: Aspects of Approximate Inference
9 Christof van Nimwegen (UU), The paradox of the guided user: assistance can be counter-effective
10 Wauter Bosma (UT), Discourse oriented summarization,
11 Vera Kartseva (VU), Designing Controls for Network Organizations: A Value-Based Approach
12 Jozsef Farkas (RUN), A Semiotically Oriented Cognitive Model of Knowledge Representation
13 Caterina Carraciolo (UVA), Topic Driven Access to Scientific Handbooks
14 Arthur van Bunningen (UT), Context-Aware Querying; Better Answers with Less Effort
15 Martijn van Otterlo (UT), The Logic of Adaptive Behavior: Knowledge Representation and Algorithms for the Markov Decision Process Framework in First-Order Domains.
16 Henriette van Vugt (VU), Embodied agents from a user's perspective
17 Martin Op 't Land (TUD), Applying Architecture and Ontology to the Splitting and Allying of Enterprises
18 Guido de Croon (UM), Adaptive Active Vision
19 Henning Rode (UT), From Document to Entity Retrieval: Improving Precision and Performance of Focused Text Search
20 Rex Arendsen (UVA), Geen bericht, goed bericht. Een onderzoek naar de effecten van de introductie van elektronisch berichtenverkeer met de overheid op de administratieve lasten van bedrijven
21 Krisztian Balog (UVA), People Search in the Enterprise
22 Henk Koning (UU), Communication of IT-Architecture
23 Stefan Visscher (UU), Bayesian network models for the management of ventilator-associated pneumonia
24 Zharko Aleksovski (VU), Using background knowledge in ontology matching
25 Geert Jonker (UU), Efficient and Equitable Exchange in Air Traffic Management Plan Repair using Spender-signed Currency
26 Marijn Huijbregts (UT), Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled
27 Hubert Vogten (OU), Design and Implementation Strategies for IMS Learning Design
28 Ildiko Flesch (RUN), On the Use of Independence Relations in Bayesian Networks
29 Dennis Reidsma (UT), Annotations and Subjective Machines - Of Annotators, Embodied Agents, Users, and Other Humans
30 Wouter van Atteveldt (VU), Semantic Network Analysis: Techniques for Extracting, Representing and Querying Media Content
31 Loes Braun (UM), Pro-Active Medical Information Retrieval
32 Trung H. Bui (UT), Toward Affective Dialogue Management using Partially Observable Markov Decision Processes
33 Frank Terpstra (UVA), Scientific Workflow Design; theoretical and practical issues
34 Jeroen de Knijf (UU), Studies in Frequent Tree Mining
35 Ben Torben Nielsen (UvT), Dendritic morphologies: function shapes structure

2009

1 Rasa Jurgelenaite (RUN), Symmetric Causal Independence Models
2 Willem Robert van Hage (VU), Evaluating Ontology-Alignment Techniques
3 Hans Stol (UvT), A Framework for Evidence-based Policy Making Using IT
4 Josephine Nabukenya (RUN), Improving the Quality of Organisational Policy Making using Collaboration Engineering
5 Sietse Overbeek (RUN), Bridging Supply and Demand for Knowledge Intensive

Tasks - Based on Knowledge, Cognition, and Quality

6 Muhammad Subianto (UU), Understanding Classification

7 Ronald Poppe (UT), Discriminative Vision-Based Recovery and Recognition of Human Motion

8 Volker Nannen (VU), Evolutionary Agent-Based Policy Analysis in Dynamic Environments

9 Benjamin Kanagwa (RUN), Design, Discovery and Construction of Service-oriented Systems

10 Jan Wielemaker (UVA), Logic programming for knowledge-intensive interactive applications

11 Alexander Boer (UVA), Legal Theory, Sources of Law & the Semantic Web

12 Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin), perating Guidelines for Services

13 Steven de Jong (UM), Fairness in Multi-Agent Systems

14 Maksym Korotkiy (VU), From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)

15 Rinke Hoekstra (UVA), Ontology Representation - Design Patterns and Ontologies that Make Sense

16 Fritz Reul (UvT), New Architectures in Computer Chess

17 Laurens van der Maaten (UvT), Feature Extraction from Visual Data

18 Fabian Groffen (CWI), Armada, An Evolving Database System

19 Valentin Robu (CWI), Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets

20 Bob van der Vecht (UU), Adjustable Autonomy: Controling Influences on Decision Making

21 Stijn Vanderlooy (UM), Ranking and Reliable Classification

22 Pavel Serdyukov (UT), Search For Expertise: Going beyond direct evidence

23 Peter Hofgesang (VU), Modelling Web Usage in a Changing Environment

24 Annerieke Heuvelink (VUA), Cognitive Models for Training Simulations

25 Alex van Ballegooij (CWI), "RAM: Array Database Management through Relational Mapping"

26 Fernando Koch (UU), An Agent-Based Model for the Development of Intelligent Mobile Services

27 Christian Glahn (OU), Contextual Support of social Engagement and Reflection on the Web

28 Sander Evers (UT), Sensor Data Management with Probabilistic Models

29 Stanislav Pokraev (UT), Model-Driven Semantic Integration of Service-Oriented Applications

30 Marcin Zukowski (CWI), Balancing vectorized query execution with bandwidth-optimized storage

31 Sofiya Katrenko (UVA), A Closer Look at Learning Relations from Text

32 Rik Farenhorst (VU) and Remco de Boer (VU), Architectural Knowledge Management: Supporting Architects and Auditors

33 Khiet Truong (UT), How Does Real Affect Affect Affect Recognition In Speech?

34 Inge van de Weerd (UU), Advancing in Software Product Management: An Incremental Method Engineering Approach

35 Wouter Koelewijn (UL), Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling

36 Marco Kalz (OUN), Placement Support for Learners in Learning Networks

37 Hendrik Drachsler (OUN), Navigation Support for Learners in Informal Learning Networks

38 Riina Vuorikari (OU), Tags and self-organisation: a metadata ecology for learning resources in a multilingual context

39 Christian Stahl (TUE, Humboldt-Universitaet zu Berlin), Service Substitution – A Behavioral Approach Based on Petri Nets

40 Stephan Raaijmakers (UvT), Multinomial Language Learning: Investigations into the Geometry of Language

41 Igor Berezhnyy (UvT), Digital Analysis of Paintings

42 Toine Bogers, Recommender Systems for Social Bookmarking

43 Virginia Nunes Leal Franqueira (UT), Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients

44 Roberto Santana Tapia (UT), Assessing Business-IT Alignment in Networked Organizations

45 Jilles Vreeken (UU), Making Pattern Mining Useful

46 Loredana Afanasiev (UvA), Querying XML: Benchmarks and Recursion

2010

1 Matthijs van Leeuwen (UU), Patterns that Matter

2 Ingo Wassink (UT), Work flows in Life Science
3 Joost Geurts (CWI), A Document Engineering Model and Processing Framework for Multimedia documents
4 Olga Kulyk (UT), Do You Know What I Know? Situational Awareness of Colocated Teams in Multidisplay Environments
5 Claudia Hauff (UT), Predicting the Effectiveness of Queries and Retrieval Systems
6 Sander Bakkes (UvT), Rapid Adaptation of Video Game AI
7 Wim Fikkert (UT), Gesture interaction at a Distance
8 Krzysztof Siewicz (UL), Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments
9 Hugo Kielman (UL), A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging
10 Rebecca Ong (UL), Mobile Communication and Protection of Children
11 Adriaan Ter Mors (TUD), The world according to MARP: Multi-Agent Route Planning
12 Susan van den Braak (UU), Sensemaking software for crime analysis
13 Gianluigi Folino (RUN), High Performance Data Mining using Bio-inspired techniques
14 Sander van Splunter (VU), Automated Web Service Reconfiguration
15 Lianne Bodenstaff (UT), Managing Dependency Relations in Inter-Organizational Models
16 Sicco Verwer (TUD), Efficient Identification of Timed Automata, theory and practice
17 Spyros Kotoulas (VU), Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications
18 Charlotte Gerritsen (VU), Caught in the Act: Investigating Crime by Agent-Based Simulation
19 Henriette Cramer (UvA), People's Responses to Autonomous and Adaptive Systems
20 Ivo Swartjes (UT), Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative
21 Harold van Heerde (UT), Privacy-aware data management by means of data degradation
22 Michiel Hildebrand (CWI), End-user Support for Access to Heterogeneous Linked Data
23 Bas Steunebrink (UU), The Logical Structure of Emotions
24 Dmytro Tykhonov, Designing Generic and Efficient Negotiation Strategies
25 Zulfiqar Ali Memon (VU), Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective
26 Ying Zhang (CWI), XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines
27 Marten Voulon (UL), Automatisch contracteren
28 Arne Koopman (UU), Characteristic Relational Patterns
29 Stratos Idreos(CWI), Database Cracking: Towards Auto-tuning Database Kernels
30 Marieke van Erp (UvT), Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval
31 Victor de Boer (UVA), Ontology Enrichment from Heterogeneous Sources on the Web
32 Marcel Hiel (UvT), An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems
33 Robin Aly (UT), Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval
34 Teduh Dirgahayu (UT), Interaction Design in Service Compositions
35 Dolf Trieschnigg (UT), Proof of Concept: Concept-based Biomedical Information Retrieval
36 Jose Janssen (OU), Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification
37 Niels Lohmann (TUE), Correctness of services and their composition
38 Dirk Fahland (TUE), From Scenarios to components
39 Ghazanfar Farooq Siddiqui (VU), Integrative modeling of emotions in virtual agents
40 Mark van Assem (VU), Converting and Integrating Vocabularies for the Semantic Web
41 Guillaume Chaslot (UM), Monte-Carlo Tree Search
42 Sybren de Kinderen (VU), Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach
43 Peter van Kranenburg (UU), A Computational Approach to Content-Based Retrieval of Folk Song Melodies
44 Pieter Bellekens (TUE), An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain
45 Vasilios Andrikopoulos (UvT), A theory and model for the evolution of software services

46 Vincent Pijpers (VU), e3alignment: Exploring Inter-Organizational Business-ICT Alignment

47 Chen Li (UT), Mining Process Model Variants: Challenges, Techniques, Examples

48 Milan Lovric (EUR), Behavioral Finance and Agent-Based Artificial Markets

49 Jahn-Takeshi Saito (UM), Solving difficult game positions

50 Bouke Huurnink (UVA), Search in Audiovisual Broadcast Archives