Understanding and supporting information seeking tasks across multiple sources

Amin, A.K.

Citation for published version (APA):
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

<table>
<thead>
<tr>
<th>Research</th>
<th>Fact Finding</th>
<th>Information Gathering</th>
<th>Non-Goal Oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morrison et al. (2001)</td>
<td>find</td>
<td>compare, collect</td>
<td>monitor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>understand, explore</td>
<td></td>
</tr>
<tr>
<td>Sellen et al. (2002)</td>
<td>finding</td>
<td>information gathering</td>
<td>browsing</td>
</tr>
<tr>
<td>Marchonini et. al. (2006)</td>
<td>look up</td>
<td>learn, investigate</td>
<td>-</td>
</tr>
<tr>
<td>Kellar et al. (2007)</td>
<td>fact finding</td>
<td>information gathering</td>
<td>just browsing</td>
</tr>
</tbody>
</table>

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?
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
1.5 Introduction

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 identifying 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.


Additional publications by the author are:


