Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

INS

Information Systems

INformation Systems

Understanding experts' information seeking needs: a user study in the cultural heritage domain

A.K. Amin, J.R. van Ossenbruggen, L. Hardman, A. van Nispen

REPORT INS-E0707 SEPTEMBER 2007

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3681

Understanding experts' information seeking needs: a user study in the cultural heritage domain

ABSTRACT
We report on our user study on the information seeking behavior of cultural heritage experts and the sources they use to carry out search tasks. Seventeen experts from nine cultural heritage institutes in the Netherlands were interviewed and asked to answer questionnaires about their daily search activities. The interviews helped us to better understand their search motivations, types, sources and tools. A key finding of our study is that the majority of search tasks involve relatively complex information gathering. This is in contrast to the relatively simple fact-finding oriented support provided by current tools. We describe a number of strategies that experts have developed to overcome the inadequacies of their tools. Based on an analysis of these strategies and the information needs of the participants, we derive a list of functional and interface recommendations for future search tools designed explicitly to support their search tasks.

2000 Mathematics Subject Classification: -
1998 ACM Computing Classification System: H.3.3 [Information Search and Retrieval]; H.5 [H.5.2 User Interfaces];
Keywords and Phrases: cultural heritage experts, user study, search behavior, task analysis

Understanding Experts' Information Seeking Needs

A User Study in the Cultural Heritage Domain

ABSTRACT

We report on our user study on the information seeking behavior of cultural heritage experts and the sources they use to carry out search tasks. Seventeen experts from nine cultural heritage institutes in the Netherlands were interviewed and asked to answer questionnaires about their daily search activities. The interviews helped us to better understand their search motivations, types, sources and tools. A key finding of our study is that the majority of search tasks involve relatively complex information gathering. This is in contrast to the relatively simple fact-finding oriented support provided by current tools. We describe a number of strategies that experts have developed to overcome the inadequacies of their tools. Based on an analysis of these strategies and the information needs of the participants, we derive a list of functional and interface recommendations for future search tools designed explicitly to support their search tasks.

Author Keywords

cultural heritage experts, user study, search behavior, task analysis

ACM Classification Keywords

H.3.3 [Information Search and Retrieval]; H.5 [H.5.2 User Interfaces];

INTRODUCTION

This study was motivated by the need to understand the information seeking behavior of cultural heritage experts and to find ways of improving their search activity. Searching for cultural heritage information 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). Most research in this area has so far focused on non-expert users, such as museum visitors. Even though we focus this study on experts in the cultural heritage domain, the issues raised in our study are relevant and similar to many other knowledge-intensive domains.

Key findings include, first, that the daily search tasks are dominated by a range of different (relatively high level) information gathering tasks, while the tools tend to be geared towards support for (lower level) fact finding tasks. Second, we found that for complex fact finding tasks, experts use the well-known strategy of starting with a simple query, followed by iterative refinement of the query into a more complex query. Few of their tools, however, support the user in this incremental building of complex queries, even though most tools provide both "basic" and "advanced" query forms. Third, we found that many search tasks require experts to use and combine results from multiple sources, while tool support is typically geared towards a single source. In summary, our study suggests that current search applications do not support experts' information seeking tasks. As a result, experts do not use search applications optimally and compensate by doing things manually or using other means to get information, such as simply asking other colleagues.

The paper 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 the current tools support the experts' search tasks and illustrate inadequate tool support with concrete examples given by experts during the interviews. Finally, we propose a number of functional and interface designs with examples that would improve the cultural heritage search tools.

RELATED WORK

The motivations behind searching information have been studied extensively [2, 3, 5, 18, 19]. Research by Broeder [2], extended by Rose et al. [19], 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. [5] 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 [4].
Kellar et al. [11] compared previous research on information seeking task categories [5, 17, 20, 21] 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 [10, 11, 12], see Figure 1):

1. Fact Finding: users ask goal oriented and focused ques-

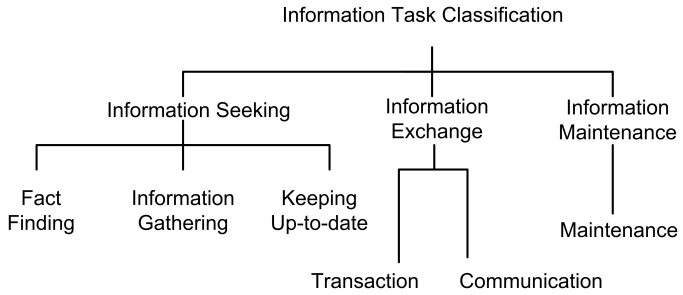

Figure 1. Classification of Information Tasks, adapted from [10]

tions; 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. Keeping Up-to-date[1]: 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.

Expert and novice users tend to display different information seeking behavior. Hoelscher et al. [8] looked at the differences in search strategies between four user groups: domain (economics) experts/novices and technology (web) expert/novices. They suggest that domain experts tend to use domain-specific terminology as one of the strategies to search, while web experts tend to use more advanced query constructs, such as `and` or "+". Jenkins et al. [9] investigated if there are web seeking pattern differences between domain (nursing) experts/novices and technology (web) experts/novices. They found out that there are differences in the breadth and depth of scanning behavior when searching. These studies show that there is a difference in search behavior for expert and novice users observed in different domains. We found, however, no prior research that studied experts' search behavior in the cultural heritage domain.

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]We use the term *Keeping up-to-date* rather then the original term *Browsing* proposed by Kellar et al., to reflect the breadth of this task.

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.

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 where 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.

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 have captured key aspects of the cultural heritage experts' information seeking behavior.

Table 1. Participants' Demography (total:17 people)

	Experience with digital libraries:	expert: 3, intermediate: 12, basic: 2 *)
	Expert role: **)	researcher: 6, curator: 8, registrar: 3, IT: 3, teacher: 2, student: 3
	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
	Affiliation: **)	museum: 8, company/freelance: 2, university: 5, cultural heritage institution: 8
*)	expert:	IT experts, also manage the museum information systems.
	intermediate:	Intensive use of search engines and online and offline digital libraries.
	basic:	Minimal Internet/computer usage, e.g only email and Word. Prefer to use traditional libraries.
**)	A single expert can have more then one role and/or affiliation.	

Participants

Based on recommendations from ICN[2], we recruited experts who frequently search for information related to cultural heritage (see Table 1). In total, 17 professionals participated from 9 different cultural heritage institutions in the Netherlands (five museums, two companies, universities and ICN itself). 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 four 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 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 help out the researchers, curators and registrars work while doing their internship.

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[3]. 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 κ = 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.

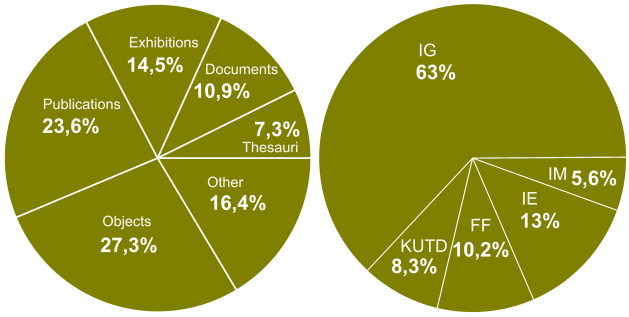

Figure 2. Types of experts' use cases (left).
Types of experts' information seeking tasks (right).

USER STUDY RESULTS

This section is divided into three parts, reflecting the three research questions of 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.

Why do experts search?

We classified all use cases into five groups (see also the left of Figure 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]The Netherlands Institute for Cultural Heritage `http://www.icn.nl/`

[3]Full use case description available from `http://e-culture.multimedian.nl/user-studies/2007-Feb/`

Table 2. Frequency mention of source (total:204 mentions)

	Source	
1	Literature	21.4%
2	Archives and catalogs	18.9%
3	Personal contact	10.2%
4	Visit exhibition/museum	1.5%
	Total offline sources:	**52.0%**
5	Reputable website	18.9%
6	Collection management system	13.8%
7	Search engine	8.7%
8	Other digital sources	6.6%
	Total digital sources:	**48.0%**

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 an 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 collect terms important for the field from selected sources such as literature, dictionaries, library archives and object descriptions for 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[4] thesaurus for Dutch ethnography and the AAT/AATNED[5] for general art and architecture terms.

Where do experts search?

Experts consult a large variety of sources to look for the answers they need (see Table 2 and 3 for detail). 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[6] or Picarta[7]. 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. E.g. communication between 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[8], RKD online, galleries.nl), global gazetteer[9] and the CIA fact book[10].
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 requests process. Examples of commonly used commercial systems are TMS[11] and ADLIB[12].
7. 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 shows a summary of the sources used by the participants. Figure 3 shows the interfaces of these sources.

[4] http://www.svcn.nl/
[5] http://www.getty.edu/research/conducting_research/vocabularies/aat/
[6] http://www.rkd.nl/rkddb/
[7] http://www.picarta.nl/
[8] http://www.svcn.nl/
[9] http://www.allm-geodata.com/products1.htm
[10] https://www.cia.gov/library/publications/the-world-factbook/
[11] http://www.gallerysystems.com/products/tms.html
[12] http://www.adlibsoft.com/

Table 3. Distribution of different sources usage for different types of use cases (total:204 items)

Use case type vs. Source	Books		Rep.Website		Archive		Search Eng		Museum DB		Person		Other		Visit Exh.Museums	
1.Objects	9	19.6%	9	23.7%	11	28.2%	3	17.6%	8	28.6%	4	20%	1	7.7%	1	33.3%
2.Exhibitions	9	19.6%	7	18.4%	10	25.6%	4	23.5%	3	10.7%	3	15%	1	7.7%	1	33.3%
3.Publications	17	37%	10	26.3%	8	20.5%	4	23.5%	7	25%	6	30%	8	61.5%	1	33.3%
4.Documentation	4	8.7%	2	5.3%	2	5.1%	1	5.9%	7	25%	0	0%	0	0%	0	0%
5.Thesauri	4	8.7%	7	18.4%	1	2.6%	2	11.8%	2	7.1%	1	5%	1	7.7%	0	0%
6.Other	3	6.5%	3	7.9%	7	17.9%	3	17.6%	1	3.6%	6	30%	2	15.4%	0	0%
Total	46		38		39		17		28		20		13		3	

Table 4. Distribution of the experts' information seeking tasks (total:110 use cases)

Search task vs. Use case type		Total	Thesauri		Documents		Exhibitions		Publications		Objects		Other	
Fact finding (FF)	10.2%	11	0	0%	1	8%	0	0%	5	16%	3	12%	2	13%
Information Gathering (IG)	63.0%													
1.Comparison		9	0	0%	2	17%	0	0%	2	7%	3	12%	2	13%
2.Relationship search		6	0	0%	0	0%	2	10%	1	3%	1	4%	2	13%
3.Topic search		39	4	57%	4	33%	7	35%	10	32%	8	32%	6	40%
4.Combination		9	0	0%	1	8%	3	15%	3	10%	2	8%	0	0%
5.Exploration		7	0	0%	0	0%	5	25%	1	3%	1	4%	0	0%
Keeping Up-to-date (KUTD)	8.3%													
1.Active		7	1	14%	1	8%	0	0%	1	3%	4	16%	0	0%
2.Passive		2	0	0%	0	0%	0	0%	1	3%	0	0%	1	7%
Information Management (IM)	5.6%	6	0	0%	2	17%	1	5%	2	7%	1	4%	0	0%
Information Exchange (IE)	13.0%													
1.Transaction		2	0	0%	1	8%	1	5%	0	0%	0	0%	0	0%
2.Communication		12	2	28%	0	0%	1	5%	5	16%	2	8%	2	13%

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.

What are the experts' search tasks?

We classified the use cases into different tasks according to the Information Task Classification. The result is shown in the right of Figure 2. The breakdown of each task in subtasks is summarized in Table 4. In the remainder of this section 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.

Fact Finding

Fact Finding search questions vary from simple to very complex. Typical examples of simple queries include:
"What is the contact information of the gallery?" [P4]
"To which tribe/culture does this object belong?" [P1]
"From where did 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]*

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 subtasks: (See also Table 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, search 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 assciated 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 them? 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 though 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]*

Keeping Up-to-date

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.

1. Active: Going to the sources and scan trough 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-list to receive information on developments such as new exhibition announcements or reviews on new books and other publications.

DISCUSSION

The results in the previous section allowed us to identify the experts' most important information seeking tasks. In this section, we discuss the individual observations of the experts categorised by information seeking task and distill a number of problems with current tool support. We also provide illustrative examples of the strategies experts have developed to work round the inadequacies of their tools.

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.

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 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 for 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*[13] *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 to remember 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

[13] Aceh is a name of a region in Sumatra, Indonesia

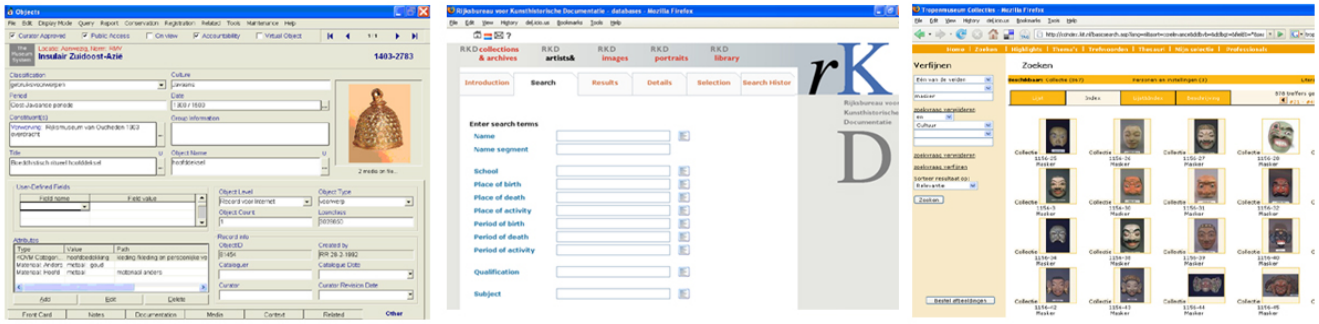

Figure 3. Interface screen shots from different types of sources, giving an impression of the typical complexity of the search interface: TMS collection management system (left), RKD online library (middle), Tropenmuseum's public website (right)

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 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 there 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 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 explorative 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.

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.

Multiple Sources Usage

In most use cases, experts need to consult multiple sources to obtain their answers (see Table 3). 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 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.

"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]*

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. Support for similar functionality is already available in other domains or proposed in the research literature. We identify several existing solutions that can be applied to solve a number of the information seeking problems in the cultural heritage domain. Other problems have no readily available solutions and so we discuss potential research directions.

Fact Finding

Our results show that experts would benefit from explicit support for formulating complex queries. Several existing interfaces could facilitate this task. *Facet browsing* [7, 27] 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. Facet browsing works best for visual resources described by structured data [27]. This interface is suitable for this task because many of the sources used by cultural heritage experts are visual, such as photographs of museum objects, visual archives and photos of artists, and stored in structured museum databases. Another query formulation interface is the *QA interface* [14, 22], such as Ask.com [14] and AnswerBus [15]. A QA interface allows the user to specify questions with the help of Natural Language Processing. The user can express their fact finding questions in natural language. Hearst [1] mentions visual interfaces, such as *VQuery, Filter-flow* and *Magic-lens*, as possible alternatives for formulating query. Hearst also mentions that visual query formulation is discovered, interestingly, more effective and faster then command language base when users need to formulate boolean query syntax, such as query with `and` and `or`. Further research is needed to find out which query formulation interfaces are most suitable for the cultural heritage domain.

Information Gathering - Comparison

Experts would benefit from a comparison interface that would allow them to see the differences and similarities of properties belonging to several objects. Comparison interfaces are used to some extent in the e-commerce domain. In an e-shop, customers often need to compare specifications of different products. Examples of e-commerce sites that include comparison interfaces are the Nokia website[16] and the Cnet product review website[17]. The interaction is in two steps: a customer makes a selection of several products which are of interest, and then the system displays the selected product in a way that that makes it easy to compare the values of the same property between different objects. Museum object data are similar to product data because they are both structured. Thus, using the same type of interface, cultural heritage experts can easily compare the values of the properties between different (sets of) museum objects.

Information Gathering - Relationship Search

To provide the relationship search required by experts, the sources used need to be structured and linked. Some of the sources that experts use are linked to each other, providing relationships among museum objects, such as between artists and art styles, or among artists. Examples of relationship interfaces among people can be found in community applications. Friendster[18] supports the *"How you are connected"* functionality, which lets a user see how s/he is connected to another person. Another example is the *DBpedia relationship finder*[19], which finds relationships between two *things* from Wikipedia. The *e-culture relation search*[20] allows users to find relationships between two objects in the cultural heritage domain. Both interfaces order the relationship results from the shortest (direct) to the longest (indirect) relationships. Finding relationships between objects is an ongoing research area, where both functionality and interface issues have yet to be solved.

Information Gathering - Topic Search

When experts are searching for as much information as possible centered around a particular topic, they need not only support for finding information related to a single term but also suggestions from the system for closely related terms, e.g. nearby geographical locations or closely related cultures. Recommender systems suggest objects that are po-

[14] `http://www.ask.com/`
[15] `http://www.answerbus.com/`
[16] `http://europe.nokia.com/products`
[17] `http://reviews.cnet.com/`
[18] `http://www.friendster.com/`
[19] `http://wikipedia.aksw.org/relfinder/`
[20] `http://e-culture.multimedian.nl/demo/path`

tentially related to a selection of objects. These have been used to enhance topic search in digital library system [16, 23]. Topic recommender for the cultural heritage domain can adopt potentially the same approach.

Information Gathering - Exploration
When experts are in search of inspiration they require an interface that allows them to broaden the scope of what they are searching for, and support them in following different trains of thought. An exploration search interface is a relatively new paradigm [6, 15]. Koshman [13] suggests that information visualization has the potential to support exploration by showing relationships between information in a novel way. White et al.[25] give an overview of possible desirable features to support exploration. While this problem has gained a lot of interest from multidisciplinary researchers [24, 26], this search activity is not yet fully understood and there is as yet little consensus on how to best support this type of search.

Keeping Up-to-date
While experts need to keep up-to-date with the development of particular topics, they should not be inundated with unwanted excesses of information. User interfaces to support keeping up-to-date, such as RSS readers, already exist. Since only few of our experts have used this type of technology, we cannot draw any conclusions about the applicability of these tools to the cultural heritage domain. The technology is relatively new and over time more experts in this area may start using it. 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. This would allow experts to view only the most interesting topics and decrease the information overload. Further research is needed to explore to what extent keeping up-to-date functionalities can and should be included in an integrated search interface for experts.

Multiple Sources Usage
Experts require access to information from multiple sources. To satisfy fact-finding requests, existing approaches that provide a single search interface to different underlying information systems, are already a big step forward. They prevent users from having to enter similar queries repeatedly, as they are currently required to by isolated tools. For most of the information gathering tasks, however, we feel such approaches are insufficient. This is because information gathering 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. Another important issue mentioned by our experts in this context is source credibility: when systems use information from different sources, source credibility needs to be taken into account at both a functional and interface level. For this, research is needed into credibility measures, visualizations helping users to assess credibility and incorporation of credibility measures in the search algorithms themselves.

CONCLUSIONS AND FUTURE WORK

We have presented the results of our study on the information seeking needs of cultural heritage experts. The goal of our 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 search applications usually support search in one particular source. Finally, we discussed potential design implications of our findings and made suggestions that may lead to better support for experts in their different search tasks.

Our next step is to concentrate on each of the tasks and see how to improve the search experience. We would like to focus our attention mostly on interfaces that could support Information Gathering, since this is our experts' main search activity but seems to have the least support. We need to 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 need to evaluate our proposed interface and to verify that we have indeed helped our experts in their search.

ACKNOWLEDGMENTS

We thank the participants from Amsterdam Tropenmuseum, Dutch National Museum of Ethnology Leiden, Rijksmuseum Amsterdam, Afrika Museum Berg en Dal, Jewish Historical Museum Amsterdam, DEN and ICN for for taking the time from their busy schedules to share their experiences with us. (left blank for anonymous review)

REFERENCES

1. R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

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

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

4. C. W. Choo, B. Detlor, and D. Turnbull. 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*, 1998.

5. C. W. Choo, B. Detlor, and D. Turnbull. *Web Work Information Seeking and Knowledge Work on the World Wide Web*. Kluwer Academic Publishers, 2001.

6. M. Hearst, A. Elliott, J. English, R. Sinha, K. Swearingen, and K.-P. Yee. Finding the flow in web site search. *Commun. ACM*, 45(9):42–49, 2002.

7. M. Hildebrand, J. van Ossenbruggen, and L. Hardman. /facet: A Browser for Heterogeneous Semantic Web Repositories. In *The Semantic Web - ISWC 2006*, pages 272–285, November 2006.

8. C. Holscher and G. Strube. Web search behavior of internet experts and newbies. *Computer Networks*, 33(1):337–346, 2000.

9. C. Jenkins, C. L. Corritore, and S. Wiedenbeck. Patterns of information seeking on the web: a qualitative study of domain expertese and web expertese. *IT&Society*, 1:64–89, 2003.

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

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

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

13. S. Koshman. Exploratory search visualization: Identifying factors affecting evaluation. EESS 2006 Workshop proceedings, https://research.microsoft.com/ ryenw/eess/index.html, 2006.

14. J. Lin, D. Quan, V. Sinha, K. Bakshi, D. Huynh, B. Katz, and D. R. Karger. The role of context in question answering systems. In *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*, pages 1006–1007, New York, NY, USA, 2003. ACM Press.

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

16. S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl. On the recommending of citations for research papers. In *CSCW '02: Proc. of the 2002 ACM conference on Computer supported cooperative work*, pages 116–125, New York, NY, USA, 2002. ACM Press.

17. J. B. Morrison, P. Pirolli, and S. K. Card. 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*, pages 163–164, New York, USA, 2001. ACM Press.

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

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

20. H. D. Rozanski, G. Bollman, and M. Lipman. Seize the occasion! the seven-segment system for online marketing, 2001.

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

22. M. Sun and J. Y. Chai. Towards intelligent QA interfaces: discourse processing for context questions. In *IUI '06: Proc. of the 11th international conference on Intelligent user interfaces*, pages 163–170, New York, NY, USA, 2006. ACM Press.

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

24. R. W. White, S. M. Drucker, G. Marchionini, M. Hearst, and m. c. schraefel. Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. In *CHI '07 extended abstracts on Human factors in computing systems*, pages 2877–2880, New York, NY, USA, 2007.

25. R. W. White, B. Kules, S. M. Drucker, and m.c. schraefel. Introduction to a special issue on supporting exploratory search. *Commun. ACM*, 49(4):36–39, 2006.

26. R. W. White, G. Muresan, and G. Marchionini. Report on the ACM SIGIR 2006 workshop on evaluating exploratory search systems. *SIGIR Forum*, 40(2):52–60, 2006.

27. K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted Metadata for Image Search and Browsing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408, Ft. Lauderdale, Florida, USA, 2003. ACM Press.