Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

INS

Information Systems

INformation Systems

Searching in the cultural heritage domain: capturing cultural heritage expert information seeking needs

A.K. Amin, L. Hardman, J.R. v. Ossenbruggen

REPORT INS-R0701 MAY 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

Searching in the cultural heritage domain: capturing cultural heritage expert information seeking needs

ABSTRACT
We report the results of a user study that captures knowledge on how cultural heritage experts search for information. We use a qualitative study technique with participants from four cultural heritage institutions in the Netherlands who were interviewed and asked to answer questionnaires about their daily work. Our goal is to acquire knowledge of their information seeking needs and the information sources they use. The paper provides an analysis and discussion of the issues that experts frequently face when searching for information.

1998 ACM Computing Classification System: H.3.7 [Digital Libraries]: user issues; H.1.2 [User/Machine Systems]: human factors
Keywords and Phrases: cultural heritage, expert, interview, user study, knowledge acquisition, information seeking task
Note: This research was supported by the MultimediaN project funded through the Bsik programme of the Dutch Government and by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content K-Space.

Searching in the Cultural Heritage Domain

Capturing cultural heritage expert information seeking needs

Alia Amin

Lynda Hardman*

Jacco van Ossenbruggen

Centrum voor Wiskunde en Informatica
Kruislaan 413C
Amsterdam, 1098SJ, The Netherlands
firstname.lastname@cwi.nl

ABSTRACT

We report the results of a user study that captures knowledge on how cultural heritage experts search for information. We use a qualitative study technique with participants from four cultural heritage institutions in the Netherlands who were interviewed and asked to answer questionnaires about their daily work. Our goal is to acquire knowledge of their information seeking needs and the information sources they use. The paper provides an analysis and discussion of the issues that experts frequently face when searching for information.

Categories and Subject Descriptors

H.3.7 [**Digital Libraries**]: user issues; H.1.2 [**User/Machine Systems**]: human factors

General Terms

Human Factors

Keywords

cultural heritage, expert, interview, user study, knowledge acquisition, information seeking task

1. INTRODUCTION

We have performed a user study and analyzed the information seeking needs of experts from the domain of cultural heritage. The study was carried out in the context of the Dutch national MultimediaN E-Culture project, which aims to develop novel tools to improve search across multiple cultural heritage collections. The original motivation for the research was thus very pragmatic: to be able to design the right system with the right interface, we needed a better understanding of the needs of the end-users in the project. We feel, however, that the results of the study are relevant for a wider audience. First, cultural heritage experts tend to search for information in a wide range of isolated and heterogeneous information sources. Moving away from closed, isolated applications to a more open and integrated environment seems to be a trend in many domains, and this study may thus provide valuable insights that could also be applied in other contexts. Second, domain experts require advanced query support. We would like to know what types of functionalities are most useful to them. For example, how important are general notions of time, location and multilingual queries, and would they appreciate query expansion results. We would like to understand how experts use advanced querying and in what context.

We feel these issues are relevant and inspirational to many knowledge-intensive domains, not just cultural heritage. Additionally, the field has a long history in systematically annotating and indexing artifacts, which potentially could help improving their own search tools. Heritage could thus be an example domain that shows the practical value of extensive annotation for search. This could benefit a wide range of annotation-based search applications, specifically those based on Semantic Web technology.

The main contribution of this paper is an in-depth study of the tasks, issues and problems that cultural heritage experts face when they are searching for information. Section 2 discusses the interview technique used, as well as previous research on information seeking needs within the cultural heritage domain. Section 3 describes the user study setup. In section 4 we report on the user study results and analyze them in section 5, where we identify a number of key issues related to information seeking. In section 6 we discuss these issues in the context of our future research plans. Concluding remarks are given in section 7.

2. RELATED WORK

There are various human computer interaction techniques that can be used to acquire knowledge on user's needs, such as qualitative and quantitative questionnaires, interviews, focus groups, lab and on-site observations, survey and participatory design [5]. These different techniques differ in the amounts of time, effort, numbers of participants and the kinds of data collected. Interviewing supports gathering rich and extensive information and give some freedom for the interviewees to express their thoughts. In a semi-structured interview, the interviewer is equipped with a list of core

*Lynda Hardman is also affiliated with the Technical University of Eindhoven.

questions that act as a road map, but the depth of the interview can be very detailed depending on the interaction between the interviewer and the interviewee. Interviewing is a good method to use in an exploration stage. Valuable use cases can be acquired or derived which are hard to discover through other techniques. The downside of this method is that an interview may take relatively long to conduct and an even longer time to transcribe and analyze the recordings afterwards. Therefore this method is only suitable for a small number of participants.

There are already studies on information seeking in the cultural heritage domain. Mostly seem to be focused on the museum visitor as the prime user, e.g. research on museum visitors' user profile [8] and how to support the museum visitors [3, 6]. Marty [4] takes another perspective and reports on tools to support curators and exhibition designers. His work focuses on how to support collaboration between a curator and exhibition designers. Research on experts information seeking needs in the cultural heritage domain has not been thoroughly investigated, although cultural heritage experts are the main users of cultural heritage information systems. The study described in the paper addresses this gap and provides insights on what are cultural heritage experts needs when searching for information for their work.

3. USER STUDY SETUP

In this section, we explain the user study setup, which consists of the procedure of the study, the participants of the study and the goal of the study.

3.1 Procedure

Our participants were recruited from four musea and the Netherlands Institute for Cultural Heritage (ICN)[1]. Most of the interviews took place at the participant's working environment. Each participant was interviewed individually with a semi-structured list of questions and was asked to answer a questionnaire. After the interview, we asked them to demonstrate the tools which they used and to give some examples on how they use it. On average, the whole interview took 3 hours. The interviews were voice recorded, pictures and screen shots of the tools which they use and questionnaires were taken and analyzed. 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 which they face.

3.2 Participants

We recruited cultural heritage professionals, whom we refer to as experts, who frequently search for (digital) information. After consulting with ICN, we obtained participants who are museum employees with three different roles: curators, registrars and researchers. In total, eight professionals, aged 22 to 54 years old, participated from four cultural heritage institutions in the Netherlands (three musea and ICN). Depending on the size of the organization, one professional can fulfill one or more expert roles. In large musea, people have clearer and more specific expert roles compared with smaller musea where one person takes responsibility for several expert roles. Explanations of the different expert roles are as follow:

Researchers (RE) perform a wide range of tasks, ranging from researching conservation techniques to building an ontology for describing contemporary art. They are typically actively looking for information and spreading knowledge into the cultural heritage community. A basic motivation is to improve the understanding and expertise within the cultural heritage domain. Researchers regularly develop guidelines, recommendations, articles and books, teach and give lectures.

Curators(CU) are responsible for the management of their collection and the corresponding documentation. This extends to arranging loans, acquiring objects and planning for exhibitions.

Registrars(RA) handle the digitization process of collections in the museum database and, depending on the size of the museum, may work together with the curators in annotating collections. Together with curators, they also handle new entries and check if information is correct, and prepare reports on the museum collection status.

The use cases all apply to these experts roles. There might be more types of expert roles, but this has not been covered in our study.

3.3 Goal

We are primarily interested in how experts search for information, and narrowed down our research into three questions: First, *what are the cultural heritage experts main tasks?* We would like to know what kind of work they do and derive realistic use cases regarding information seeking. Second, *what are the cultural heritage experts' information sources?* We would like to know which tools and applications they use to support them in information seeking. Three, *what are the issues and problems which they encounter when they search for information?* These goals were used as the basis of our user study design.

4. USER STUDY RESULTS

We identified all the use cases[2], including similar use cases from different participants (duplicated use cases) and use cases which do not involve information seeking activities in the cultural heritage domain (irrelevant use cases), such as website maintenance, project management or fund raising. We removed all duplicated and irrelevant use cases. Table 1 summarizes the results, which include the use cases, the tools used and the issues identified within the use cases.

4.1 Cultural Heritage Experts Main Tasks

We classified the use cases into four groups, based on the corresponding task:

1. *Object handling.* The main information seeking activity in this group happens when experts need to gather information for restoration, acquisition, loan or sale of an individual object. For example, when restoring the color of the painting, the curator needs to perform research on the original color of the relevant part of the painting, e.g. by searching archives or auction catalogs for an image of the original color. Another example is

[1] http://www.icn.nl/

[2] The full description of the use cases are available from http://e-culture.multimedian.nl/user-studies/2007-Feb/

that before acquisition of a painting, the curator needs to research the history of the painting to see whether it matches the rest of the collection.

2. *Planning exhibition or publication.* This requires the experts to spend a great deal of effort researching, e.g. finding topics, carrying out a comparison study with previous exhibitions and publications. The main goal is to find different and interesting perspectives. Serendipity is highly valued here. In the research phase, the main task is discovering new aspects. In the selection phase, the main task is to find and match a number of objects with the theme of the exhibition.

3. *Managing the collections' documentation.* Records in the museum database are constantly updated, for example, a new object needs to be registered or information from old collections need to be updated according to predefined form. When a new object needs to be registered, experts compare annotations of similar objects, or search further in literature, websites and library archives.

4. *Building thesauri.* Thesauri are controlled vocabularies, in this case used for annotating objects for museum records. In the participants' organizations, it is important that everyone use the same terms to express the same thing in the museum records. Experts collect terms important for the field from selected sources such as literature, dictionaries, library archives and object descriptions. Information from websites with certain reputation, such as those of other musea or cultural heritage organizations, are also used. The experts need to agree upon the proposed terms before they are included in the thesaurus. Every cultural heritage branch may have a different thesaurus, examples encountered in our use cases include the SVCN[3] thesaurus for the Dutch ethnography field or (the Dutch translation of) the AAT for general art and architecture terms.

4.2 Cultural Heritage Experts Main Sources

There is a large number and variety of sources that experts use when they look for the answers they need. These sources are:

1. Literature: magazines, dictionaries, books, publications, institute and other archives, biographies, encyclopedia, and intranet databases, and the Netherlands Institute of Art History (RKD)[4] library which has a rich collection of personal, visual and press documentation. To find the correct literature, experts also use online library portals such as Picarta[5].

2. Museum information systems: each museum typically has its own information system, which not only stores the records of all objects but is an integrated system to help all museum employees do many things for their daily work. Systems encountered in our study include The Museum System (TMS)[6], ADLIB[7] and CIS (an in house database system).

3. Notes: exhibition catalogs, object inventory cards and remarks fields made by curators.

4. Reputable websites: our participants stress that annotations for objects in their museum collection should only be taken from reliable sources. The museum curators, who are mainly responsible for the annotation form a consensus of which resource can be referenced, including online sources. Example websites which are deemed reliable are: museum websites, institution websites e.g. SVCN website, RKD, `galleries.nl`, gazetteers[8] and the CIA fact book[9] for geographical names.

5. General web search engines: Most experts whom we interviewed use Google as their search engine. In contrast with how experts use reputable websites, experts use Google to seek inspiration, especially in research to find themes and ideas for exhibitions.

5. ANALYSIS OF THE ISSUES FOUND

In this section we report our observations on a number of the issues the experts face during their work. Every issue includes discussions, including some comments from the experts themselves and the problem these experts experienced. The issues were collected from the use cases (see Table 1). To simplify the discussion, they are grouped into several main categories: integration, query semantics, query formulation, credibility and annotation.

5.1 Integration

Integration are issues related to dealing with different systems and data.

a. Dealing with many systems — Experts must deal with many different tools, online and offline, to access different databases (shown in section 4.2). The more intensive information searching the experts need to do, for example preparing an exhibition, the more diverse tools for different sources the experts use. The problem here is efficiency of work: because the systems are not connected to each other, the experts need to manually find and collect the information from many different sources.

b. Dealing with heterogeneous data — As well as having to use different systems, experts need to deal with heterogeneous data. This is partially true when looking for inspiration or conducting research: experts need to search for information and connections between objects, artist names, institutions, galleries, history and literature. Most services concentrate on providing certain information e.g. musea provide information surrounding objects, RKD Artist[10] provides information about Dutch artist database. To date, there is no known professional system that allows experts to simultaneously access information and to access relationships between heterogenous data. This is done manually which often requires extensive work.

[3] `http://www.svcn.nl/`
[4] `http://website.rkd.nl/`
[5] `http://www.picarta.nl/`
[6] `http://www.gallerysystems.com/`
[7] `http://www.adlibsoft.com/`
[8] `http://www.gazeteer.com/`
[9] `https://www.cia.gov/cia/publications/factbook/`
[10] `http://www.rkd.nl/rkddb/search.aspx`

Table 1: Summary of use cases organized by tasks for Curators (CU), Registrars (RA), and Researchers (RE)

Use cases (type of user)	Tools	Issue found
Group A. Object handling		
1. Research for object acquisition or release (CU)	RKD, Google, lexicon exhibition catalog, museum database	a.Deal with many systems b.Deal with heterogeneous data
2. Research before object restoration (CU)	old archive, old auction catalog RKD library, museum database	
3. Assessing the value of an object (CU)	RKD, Google, Lexicon	m. Navigational search
Group B. Planning exhibition or publication		
4. Planning (RE)		
5. Research (RE,CU)	catalogs art and history literature museum database artist, gallery website RKD online, RKD library Google, lexicon Internet, Getty database Picarta, municipal archive digital newspaper archive historical library	a. Deal with many systems f. Exploratory search m. Navigational search j. Image search i. Cultural influence and similarities n. Seek relationship e. Many language, alphabet, and spelling b. Deal with heterogeneous data n. Queries with unexpected results k. query by example
6. Comparison study with the collection (CU) of other museum	SVCN portal	
7. Object selection (CU)	museum database digital newspaper archive	g. Location query c. Narrower and broader search e. Many language, alphabet, and spelling d. Time query p. Insufficient object description
8. Documentation, presentation of research (RE)	publication presentation exhibition catalog	
Group C. Managing the collections' documentation		
9. Research on object(s) (CU)	literature certain websites library archive	a. Deal with many systems l. Query for unknown term o. Trust in service g. Location query e. Many language, alphabet, and spelling
10. Register new object(s) (RA,CU)	museum database	l. Query for unknown term o. Trust in service
11. Update existing records (RA,CU)	literature museum database certain websites	l. Query for unknown term o. Trust in service
12. Make status reports of all records (RA)	museum database	
13. Digitize and store old records (RA)	museum database	
14. Monitor the collection status (RA)	museum database	
Group D. Building Thesaurus		
15. Select, collect, and structure list of terms and quotes (RA,RE)	literature collection description certain websites library archive database	a. Deal with many systems l. Query for unknown term o. Trust in service g. Location query
16. Discussion with fellow experts for agreement (RA)	meetings,email teleconf	a. Deal with many systems
17. Disseminate knowledge to the community (RE)	publication, presentation	

5.2 Query Semantics

Query semantics are query issues which requires some degree of semantics and inference to help get better results.

c. Narrower and broader search — Experts need to use broader and narrower terms to develop an overview of the collections and see the number of hits they would get while searching in the museum database. For example, they search on the continent *Africa* to get more results of objects and search for a specific region *Cameroon* to get fewer but more specific results.

"If I want to see more objects I search on countries or large regions, but to get specific objects I type in a tribe or village name." [P1]

While some museum database support narrower and broader search, many tools do not yet support this. Many search applications work only by text string match. Experts need to manually place in all possible terms and combine them to get the results which they needed.

d. Many languages, alphabets, and spellings — When searching using a term, experts need to be aware of possible variations of a word. Objects in a collection can be annotated using different terms. The difference can be caused by:

1. Different languages to use: Annotation can be in at least one of three languages: an international language e.g. English:*island*, the official language of the museum e.g. Dutch:*eiland*, or in the original language where the object comes from e.g. Indonesian:*pulau*.
2. Different alphabet: this problem arises when the original language used is not in a Latin alphabet e.g. Arabic or Hebrew, the English translation can have multiple spellings.
3. Different spellings within a language for example: American English or British English, e.g. jewelery or jewellery

This causes problems in search, because search terms work based on word matching. Differences in spelling result in the objects not being found.

e. Exploratory search — Experts need to gather inspiration or ideas for example to prepare an exhibition, write publications, etc. We describe two examples below that illustrate how they find ideas.

Example 1 — Google is often used as a starting point when looking for inspiration. Experts would like to see what kind of information is already available out there. One of our expert [P8] described her experience looking for inspiration for an exhibition titled "The 'Jewish' Rembrandt". She used google to search and found a blog about a story told by a prominent Rabbi about how Rembrandt had a Jewish soul. This inspired her to investigate all the romantic myth that has grown up over the centuries about Rembrandt's special relationship with the Jewish people[11].

[11] http://www.jhm.nl/press.aspx?ID=6

Example 2 — Relationships between terms are often a very important inspirational source. An object or term can have many associations with other objects or terms. For example, a piece of *jewellery* can be used as an *amulet*, or it can have some *ceremonial* importance. Another example is when looking for examples of *staircase projects*, interesting related project for suggestions are on *landscapes art project* and *city planning art projects*.

"... On specific situations, (such as) in the Staircase project, I look a lot on similar examples of artwork in staircases, for instance, art projects connected to landscapes or city planning, something like that." [P4]

In the two examples above, *serendipitous discovery*, such as unexpectedly discovering new knowledge, a new perspective on the topic of interest or new relationships, is the main goal when an expert is looking for idea. The experts currently need to make these connections manually when searching. Currently, experts rely only on their cognitive and analytical skills to bring all this knowledge together when they do research. There are no tools that explicity "support" serendipitous discovery.

f. Time query — Experts want to use time as a constraint when they are searching for objects. For example, when looking for objects made during, before or after a certain time. In more complex situation, for example, a special coat was made in the ce. 17th., this objects goes under restoration as the inner padding (furring) which is made from silk has worn out. The reparation, using silk from China, is done in the ce.18th. When a curator search for all objects which were restored in the ce. 18th. she should also be able to find the coat, even though made in the ce. 17th.

"Sometimes the dates are not available in our system, it is better to search on other (search terms)." [P6]

None of the tools which the experts use to search based on time can support time queries beyond syntatic matches.

g. Location query — Names of places and regions change through time, eg. *Zaire* is now the *Democratic Republic Congo* and *Jakarta* was once called *Batavia*. Experts need to search using all possible place names and its name variations to be sure he obtains all results. This demands special geographical knowledge and a lot of research time.

In addition, borders of a region or country can change throughout time. Regions merge and dissolve into different or new countries. *Timor-Leste*, once a province named *East Timur* of the Republic of Indonesia, is a country since 2002. *Alsace*, now a part of France, historically has been swapped back and forth under German or France rule since World War I. Searching based on places or regions thus requires experts to be aware of the geopolitical changes, because their systems are not.

h. Seeking Relationships — Experts quite often search for relationships between different concepts. Examples of relationships that are interesting for experts include:

Example 1. People relationships; An expert needs to find

relationships between people.
"Does Rembrandt know any influential Jewish people?" [P8]
"Who are the heirs of the generals that participated in the Atjeh War?" [P3]
"Which gallery is a certain artist associated with?" [P4]

Example 2. Organization relationships; An expert need to find an artist which can make an art piece for the hospital's stair case.
"...I need to try clever search using google or the catalogs, I find out something about a project which leads to another project, etc. It's like detective work." [P4]

Example 3. Relationships between concepts; Sometimes the questions experts have are not yet well defined. For example, when planning an exhibition. Often experts start with a general question and work their way further to a more concrete topic. To select pieces of artwork which have a connection with a particular theme in mind.
"I look for all paintings (that is) somehow related to Amsterdam." [P6]

Relationships between people, institutions, and objects are an important search query. This type of search is not supported well in their tools, in which case, experts need to conduct further research to find relationships by manually looking into different sources.

i. Cultural influence and similarities — Cultures intertwine and influence one another, which results in similarities between subcultures. For example, artifacts which belong to the Yoruba people from Africa also exist in Brazil. These were brought by African people who came there during the colonial period. Experts know this history and need to search on location Africa and Brazil to find artifacts made by the Yoruba people.

Information cross checking between different domains, such as history and cultural heritage is needed. Sometimes it demands high knowledge expertise and a long time to research. Currently, it is done manually by the experts.

j. Image search — Experts need to search on images in different ways: using textual search and non-textual search. Textual search can be used on different conceptual levels. To search for the left image in Figure 1. a curator can search with general concepts, such as, *old man*. Another possibility is to search with specific concepts. If the user knows that the painting is depicting minister Johannes Wtenbogaert, she could use the name: *Johannes Wtenbogaert* as a more specific search term.

"Looking for images is very hard in Google, I wish they would catalog images better." [P4]

To date, finding the images through general and specific terms works only if the image has been annotated literally with these keywords (see also issue p.). More semantic matching instead of pure syntactic matching is not supported by most systems.

5.3 Query formulation

Figure 1: Left: Portrait of Johannes Wtenbogaert by Rembrandt van Rijn (1633) ©Rijksmuseum Amsterdam. Right: Finials consist of three levels, with bells in every level, shaped like a tower ©Joods Historisch Museum.

Query formulation consists of issues which requires interface solutions in addition to technologies such as image analysis and semantics to help provide results.

k. Query by example — Another way to search for image is using non-textual search, such as looking for objects with similar features, decoration, texture, etc. For example, an expert is looking for all *finials*[12] with similar shape or looking for coins with a matching emblem. In this case, shape or decoration is something which is very hard to describe textually (figure 1 right).

"For example, it is difficult to search for images of Finials (point to an image from a catalog) with the same ornament or shape as this one in any system." [P8]

Both textual and non-textual search are important when experts search for images. The main problem is that there is little support for textual image search and practically no support for non-textual image search in most commercial applications used by the cultural heritage experts.

l. Query for unknown term — Cultural heritage experts frequently face new and unfamiliar terms, whether in a foreign language or new terms, some experts use Google to quickly see what kind of result it gives back. Experts say that this is a fast way to understand the topic before pursuing further to the right literature or source.

"I use google as a dictionary, looking up new words and terms that I do not know. I use it as as a starting point. I never use google as a source though. It's more for orientating on the theme and then go for better sources." [P2]

In this case, the expert needs to quickly learn about the topic, much like looking up an encyclopedia or a dictionary. The expert uses a search engine to query an unknown term because it is faster and easier then scanning through different encyclopedias or a dictionaries. The main problem here is that search engines are not designed to be an encyclopedia or dictionary, thus experts may not always get a straightforward answer.

[12] A usually foliated ornament forming an upper extremity especially in Gothic architecture

m. Navigational search — Experts use google to point them to an artist or gallery web page.

"I usually have a method to search for 'living artist' information, I go to Google for the artist name, I try to find their own website or their gallery's website or some projects involvement which I already know about but I want to have specific images or know more about."[P4]

Much like looking up a telephone number in a telephone book, experts have a specific question in mind and they use a search engine to direct them to the website of the artist, museum or gallery. The most frequent problem is that the search engine gives back many results which may or may not be relevant.

n. Queries with unexpected results — Even though experts acknowledge the usefulness of search engines, they are also aware that the results may not always be as expected.

"It's like Russian roulette, sometimes a lot of information, eg. (Indonesian) artist information in Indonesian language, sometimes it surprises me." [P7]

Search engines typically give unexpected results because the return pages from sites in undesired languages or pages on irrelevant topics because of matching homonyms.

5.4 Credibility

Credibility is about believability, trustworthiness and perceived quality of the information presented.

o. Trust in the information source — Most experts have a number of websites which they visit regularly. The reason is there are not many websites which give reliable information. These websites are usually from musea, institutions (e.g. SVCN website), government (e.g. CIA fact book website), or libraries.

"If I have to make in the museum a description of 'Korwar'[13] those names are not in AAT but some musea have these objects, we go to the website and find descriptions, there are general information about this type of sculpture, that's how we get this information ... For the thesaurus we decided to use all the literature which we agreed upon. Several years ago we do not accept anything from Internet, but nowadays some. The curators are a little bit afraid of it, who says the information is correct?" [P2]

Experts only trust information sources which are regarded as credible. The higher the importance of the information, the more important credibility is. Credibility of the source is something which is currently often missing in the Web. Many people are not sure whether to trust the content which are presented to them when they cannot judge the authority of the source.

5.5 Annotation

Annotation is an issue that is related to the (lack of) meta data provided about the objects.

p. Insufficient object description — A large museum can have hundreds of thousands of objects. Sometimes information about the object on the system database is not complete.

"Even though I look for object in CIS (database), I still need to go to the depot and examine the objects personally there. Some things are not recorded in the database e.g. the physical condition of the object. Sometimes the information on the system is not complete." [P5]

Annotation of an object within a collection is a time consuming process which is currently done manually by registrars and curators. The problem is that many of the annotations are not complete or too limited in scope so that curators cannot find the object from the database.

6. DISCUSSION AND FUTURE WORK

Many of the issues identified above could, in theory, be solved by applying more state of the art technology and by incorporating well known results from research fields such as Information Retrieval and Human-Computer Interaction. We think there are several reasons why these issues have not been solved in the applications used by the participants of our study.

The cultural heritage community has a long and rich tradition in developing high quality vocabularies (see section 4.1). It is, therefore, surprising to see that some of the issues mentioned by the experts could be addressed by better deployment of these very vocabularies. For example, issue (c.), the lack of support for automatic query expansion could be addressed by using the explicit broader and narrower term relations that are part of virtually every vocabulary developed in the field. Spelling and other variants in geographical names (issue d. and g.) are addressed by, for example, Getty's TGN[14] and other thesauri. Spelling variants in artists' names are addressed by vocabularies such as RKD Artist and Getty's ULAN[15]. These also contain explicit interpersonal relationships between artists (issue h.). Variations in language (issue d.) should be addressed by using multilingual lexical database such as Global WordNet[16]. In addition to directly using these vocabularies and ontologies to solve simple queries, there are many benefits of using inference techniques. Using patterns of related terms, exploiting semantic different combination of patterns such as meronymy, hypernym, and hyponym in WordNet[17] for recommendation can help serendipitous discovery (issue e.)

While some individual systems might use some of these vocabularies, a more systematic deployment of vocabulary usage across the many systems is needed. The fact that most experts use a large number of independent, standalone systems, thus makes not only many search tasks time consuming (issue a.), it also hinders implementation of more effective search strategies. Typically, systems within an organization are developed by different vendors. Improving search interfaces and functionality on a per system basis by nego-

[13] Ancestor's sculpture from Papua New Guinea

[14] http://www.getty.edu/research/conducting_research/vocabularies/tgn/

[15] http://www.getty.edu/research/conducting_research/vocabularies/ulan/

[16] http://www.globalwordnet.org/

[17] http://wordnet.princeton.edu/

tiating vocabulary-based or other extensions with the many vendors involved is an unattractive, time consuming and expensive option.
Due to the heterogeneity of the data in the different systems, integrating them by a traditional data and schema integration approach is also an unattractive option. Data heterogeneity thus not only makes search tasks more difficult for the user (issue b.), it also makes implementing more integrated solutions harder for developers.

In our MultimediaN E-Culture project, we are currently researching to what extent these issues can be solved by applying Semantic Web technology. Preliminary results [2, 7] are promising. Since Semantic Web technology allows us to integrate heterogeneous data sets and multiple vocabularies in a single framework, we are optimistic that we will be able to solve many of the issues mentioned above. However, our user study also indicated key research areas we initially overlooked.
First, the public websites of most cultural heritage institutes focus on making the information about the museum's artifacts available on line. Thus, our assumption was that experts too mainly search on information about artifacts. The study proved this assumption to be false: while artifact related search tasks are indeed important, other key search targets include information about persons, galleries and more thematic searches (issue b.).
Second, our initial integrated interface to several sources paid insufficient attention to the role of data credibility and the importance of the authority of the original data provider. This study showed that it is important for experts to know where the data originates (issue o.). How to convey this credibility information in the user interface is, however, still unclear. Eysenbach and Kohler [1] describe factors that influence the credibility of Internet based information. We are currently setting up a user experiment to better understand credibility issues for Semantic Web systems and how to visualize it to maximize transparency.
Third, the study showed that in some cases, the functionality required by the expert is supported by their systems, but they do not use it because either they are not aware of the feature or they find the use of the feature too complex. Some of the earlier systems used by the cultural heritage organizations were indeed based on structured vocabularies or hierarchical, thesaurus-based interfaces, but these were replaced by systems based on flat keyword search because these are simpler and better known to the current generation of users who are already familiar with keyword search on the Web. We thus need to make sure our interfaces are sufficiently simple to be used. We are therefore currently redesigning the vocabulary-based multi-faceted interface presented in [2]. We need to evaluate the usability of our new interface based on the use cases found in this study.

7. CONCLUSION AND FUTURE WORK

In this paper, we explained our efforts in acquiring knowledge on information seeking from cultural heritage experts. We demonstrated that using a user study, we can capture real use cases and real issues that the experts experience in their daily tasks. We have discussed the cultural heritage experts' information seeking tasks and the kinds of tools that they use. We also identified issues encountered by the experts that can be categorized into five groups: integration, query semantics, query formulation, credibility and annotation. We discussed possible solutions for these issues that comprise combinations of different technologies such as integration, background knowledge, thesauri, visual analysis and interface. We acknowledge that there are some challenges with the solutions which we discuss above. This is something that needs to be explored in future research. For our next step, we plan to implement some of these solutions in our demonstrator[18] [7] and evaluate to what extent the solutions help cultural heritage experts.

8. ACKNOWLEDGMENTS

We would like to thank DEN, ICN, Joods Historisch Museum, Rijksmuseum voor Volkenkunde and the Afrika Museum for their cooperation. Marie-France van Oorsouw, Annemiek Teesing, Marco de Niet, Annelies van Nespen, Lloyd Rutledge, Lora Aroyo, the MultimediaN E-Culture project members and our colleagues at CWI provided valuable input in the design and analysis of the user study. Raphael Troncy and Zeljko Obrenovic provided valuable feedback on previous versions of this paper. Michiel Hildebrand provided technical support. We are extremely grateful to all the study participants for taking the time from their busy schedules to share their experiences with us. This research was supported by the MultimediaN project funded through the BSIK programme of the Dutch Government and by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content K-Space.

9. REFERENCES

[1] G. Eysenbach and C. Köhler. 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*, pages 573–577, March 2002.

[2] 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.

[3] P. Marti, A. Rizzo, L. Petroni, G. Tozzi, and M. Diligenti. Adapting the museum: A non-intrusive user modeling approach, 1999.

[4] P. F. Marty. Online exhibit design: The sociotechnological impact of building a museum over the world wide web. In *Journal of the American Society for Information Science*, volume 51, pages 24–32, 2000.

[5] D. J. Mayhew. *The usability engineering lifecycle: a practitioner's handbook for user interface design.* Morgan Kaufmann Publishers Inc., 1999.

[6] F. Paterno and C. Mancini. Effective levels of adaptation to different types of users in interactive museum systems. *J. Am. Soc. Inf. Sci.*, 51:5–13, 2000.

[7] G. Schreiber et al. MultimediaN E-Culture Demonstrator. In *The Semantic Web - ISWC 2006*, volume 4273 of *LNCS*, pages 951–958. Springer, 2006.

[8] E. Véron and M. Levasseur. *Ethnographie de l'exposition.* Bibliothèque Publique d'Information, Centre Georges Pompidou, Paris., 1983.

[18] `http://e-culture.multimedian.nl/demo/search`