CWI

Establishing requirements for information gathering tasks

A.K. Amin

INS-E0905

Centrum Wiskunde & Informatica

Centrum Wiskunde & 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)
Science Park 123, 1098 XG Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3681

Establishing Requirements for Information Gathering Tasks

Alia Amin

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

Abstract. This PhD project aims at understanding and supporting the complex activities of information gathering. To date, most search applications support one aspect of search namely low-level keyword-based search to find documents. However, in reality, users search tasks are often high-level search tasks, such as comparing differences between art objects. Thus, there is a mismatch between the users search needs and the available search tools. This research investigates information gathering characteristics and how alternative search interfaces could support them. Design recommendations and guidelines for interfaces to support information gathering will be designed and verified through lab experiments and evaluation in different domains.

2000 Mathematics Subject Classification: -
1998 ACM Computing Classification System: H.1.2 User/Machine Systems
Keywords and Phrases: human factors; information seeking tasks; cultural heritage

Motivation — Information gathering tasks involve collecting information, often from various sources, to fulfill a higher level goal, such as making a decision, writing a report, and completing a project [3]. To date, most search interfaces and functionalities are typically designed for *fact finding* by means of keyword matching such as the simple or advanced (Google-like) search interfaces. Nevertheless, it is not certain whether these interfaces are sufficient for information gathering tasks. This research aims at addressing the following questions: What are the characteristics of information gathering tasks? How can information gathering tasks be supported? What kinds of search interface are best suited for which information gathering tasks?
The following section will discuss three research stages. Stage 1 and 2 describe user studies on understanding information gathering behavior for different settings: cultural heritage experts in their daily professional work and visitors in a museum. These stages provide requirements to develop applications that support information gathering tasks. Based on the requirements derived from earlier studies, I explain the future work in stage 3 on design and evaluation of information gathering interfaces. Finally, this paper concludes with discussions on research challenges and the research contribution.

Stage 1: Identifying the Characteristics of Information Gathering: A User Study on the Cultural Heritage Experts
To better understand the information gathering task characteristics, I extended the work on information seeking taxonomy [3] and used interview techniques for a user study on the cultural heritage experts [1]. Typically, these experts search intensively using information sources that are rich and heterogeneous, both digital and non-digital, and which combine text as well as other media (e.g. image and video). In this study, a

semi-structured interview was conducted. Additionally, participants were also asked to demonstrate the tools they used and to give some examples on how they used them. In total, 17 experts (i.e. researchers, curators, registrars, teachers and students) were recruited from 9 different cultural heritage institutions (5 museums, 2 companies, a university and ICN). From the interviews, 110 information seeking tasks were collected. All information seeking tasks were classified independently by two reviewers into one of 5 different groups of information task categories [2]. Of the 110 information seeking tasks collected, 70 tasks (63%) were identified as information gathering search task. Within these 70 information gathering tasks, I identified subcategories, based on the similarity between the tasks within each group:

COMPARISON SEARCH — involve gathering information to compare differences and similarities between objects or sets of objects in a collection, e.g. *"For our exhibition, which objects from Aceh that are missing in our own collection can we borrow from that museum." [P3]*

RELATIONSHIP SEARCH — find relationships between individual pieces of information. For example, one expert needs to find out people related to the artist Rembrandt to create a story for the exhibition. *"I check old archives, history books, collect the names and make the connections." [P8]*

TOPIC SEARCH — queries can typically be formulated as "Tell me about" questions. For example, an art history student collects materials from museums and libraries for her thesis on African trade. *"Are there any objects in the museum related to the African trade in the 17th theme?" [P17]*

EXPLORATORY SEARCH — typically not goal directed. Instead, the expert may associatively follow one train of thought after another. For example, one expert described her experience looking for an exhibition theme for "The Jewish Rembrandt". She used Google and found a blog about a story about how Rembrandt had a Jewish Soul. This inspires her to investigate the romantic myth about Rembrandts special relationship with the Jewish people.

COMBINATION — find 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, one expert needs to find several artists with specialties that would match with her art project requirements. *"I select around 5 best artists which I think are suitable for the job, then I collect and present their portfolio to the client." [P4]*

The user study revealed that information gathering is the main task for experts (63% of the 110 information seeking tasks identified) and information gathering tasks are diverse and laborious. Most of the search tasks identified from this study have relatively high level goals and require using multiple information sources. In contrast, the tool support of these experts tends to focus on lower level fact finding tasks, using a single information source. I also observed that because current search tools do not support these information gathering tasks optimally, experts try to compensate by doing everything manually, rely on experience and knowledge from themselves and from their peers. Although this is not always possible, as one curator explains: *"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]*

Stage 2 - Work in Progress: Understanding Information Gathering in Ubiquitous Environment: A study on Museum Visitors

This section discusses the work in progress. Study in stage 1 provides insight on information gathering behavior for professionals in a static environment (i.e. on a desk, mainly working with a desktop). As a complement to stage 1 research, I am conducting stage 2 research to understand information gathering behavior in a dynamic and ubiquitous setting for general users. To narrow down the research, I focus to study on visitors of a museum. The research questions are: *What are the museum visitors information gathering tasks? Do the current tools support this need? How can we help museum visitors get the information that they would like to know?*

To capture museum visitors' micro and latent information seeking needs and information gathering tasks in a dynamic environment, a passive capture digital diary tool similar to an observational tool discussed in [2] is under development. The tool will help us capture not only users' search task activities, but also the movement, environment and context when the activities occurred. With the rich data collected, I hope the study will reveal new insights on information gathering in ubiquitous environment that supplements the stage 1 research.

Stage 3 - Future Work: Developing Information Gathering Interfaces

Stage 1 and 2 research help construct a better understanding of the users' information gathering needs and derive requirements for a better search applications. This section discusses future work plan. The aim of the research is to explore variations of interfaces that supports different types of information gathering task. The research will follow a user-centered design approach. As a first step, I intend to investigate comparison search, a search task which was identified as one important information gathering task in Stage 1. The research execution will be divided into three phases: *(1) Literature study phase*: The work is primarily on investigating previous research and existing interfaces that might help users in the comparison search task. I will also derive specific user and functional requirements, based on the study in stage 1 and 2 and related the literature. *(2) Design and implementation phase*: The aim of this phase is to come up with several variations of interface and interaction design mock ups, which will later be implemented and evaluated. *(3) Evaluation phase*: In this phase, there will be experiments where several different variants of interfaces, in particular alternatives interaction style and visualizations, will be evaluated. To answer different research questions, there will be several types of evaluation: laboratory experiment or evaluation in different domains.

LABORATORY EXPERIMENT — The laboratory experiments are designed to test the performance of users with several variants of interfaces. The data will be restricted, coming from a few information sources, e.g. several museum collections. Participants are asked to carry out specific assignments. For the different interfaces, objective measurement will be taken, such as time, score or topics covered, as well as subjective measurement, such as users confidence and satisfaction with the interface.

EVALUATION IN DIFFERENT DOMAINS — This study aims to test to what extent the new search interfaces support the same information gathering task in other domains, such as (mobile) e-commerce or (mobile) digital library.

The results from these evaluations will complement each other to provide a more comprehensive understanding of interface requirements to support comparison search task.

Research challenges — Methods for evaluating information gathering interfaces are often not straightforward. In some information gathering applications, such as exploratory search, performance based measures such as search time and error are insufficient [5]. Other objective measures such as topics covered, links followed, number of queries per session or duration of session have also been used to evaluate such applications. Alternatively, experienced-based measures such as, user confidence in completeness, exposure to different strategies, search productivity and engagement [4] might be better indicators to see whether an interface is suitable for such information gathering tasks. There is currently no consensus on how to evaluate these types of applications. Further research is required in this area to find the appropriate method of evaluation.

Research contribution — The contribution of this research lies on in-depth analysis of information gathering search tasks in different domains, design recommendations for interfaces to support information gathering tasks and evaluation of search tools that may support these tasks. We offer insights that will benefit researchers and practitioners from different communities, such as Human-Computer Interaction, Information Retrieval, Digital Library and Semantic Web with regards to the guidelines, lessons learned, and recommendations on how to support real users information gathering search task.

Acknowledgments — I thank my supervisors Lynda Hardman and Jacco van Ossenbruggen for their support and for believing in this research and Zeljko Obrenovic for great feedbacks on the paper. This research was supported by the Dutch BSIK MultimediaN project and the EC FP6-027026 KSpace.

References

1. A. Amin, L. Hardman, J. van Ossenbruggen, and A. van Nispen. Understanding cultural heritage experts' information seeking tasks. In Proc. *JCDL'08*, New York, USA, 2008. ACM Press.
2. D. Byrne, A. R. Doherty, G. J. Jones, A. Smeaton, S. Kumpulainen, and K. Jarvelin. The sensecam as a tool for task observation. In *People and Computers XXI Interaction*, 2008 (to appear).
3. M. Kellar, C. Watters, and M. Shepherd. The impact of task on the usage of web browser navigation mechanisms. In Proc. *GI'06*, Toronto, Canada, 2006. Canadian Info. Proc. Soc.
4. H. OBrien and E. G. Toms. Evaluating engagement in interactive search. In *ACM SIGIR 2007 Workshop report on Web Information Seeking and Interaction, SIGIR Forum.*, 2007.
5. R. W. White, G. Muresan, and G. Marchionini. Report on the ACM SIGIR 2006 workshop on evaluating exploratory search systems. *SIGIR Forum*, 40(2), 2006.

Centrum Wiskunde & Informatica (CWI) is the national research institute for mathematics and computer science in the Netherlands. The institute's strategy is to concentrate research on four broad, societally relevant themes: earth and life sciences, the data explosion, societal logistics and software as service.

Centrum Wiskunde & Informatica (CWI) is het nationale onderzoeksinstituut op het gebied van wiskunde en informatica. De strategie van het instituut concentreert zich op vier maatschappelijk relevante onderzoeksthema's: aard- en levenswetenschappen, de data-explosie, maatschappelijke logistiek en software als service.

Bezoekadres:
Science Park 123
Amsterdam

Postadres:
Postbus 94079, 1090 GB Amsterdam
Telefoon 020 592 93 33
Fax 020 592 41 99
info@cwi.nl
www.cwi.nl

CWI

Centrum Wiskunde & Informatica