Documenting and Studying the Use of Assigned Search Tasks: RepAST

Luanne Freund
School of Library, Archival and Information Studies
University of British Columbia (Canada)
Luanne.freund@ubc.ca

Barbara M. Wildemuth
School of Information and Library Science
University of North Carolina at Chapel Hill (USA)
wildemuth@unc.edu

ABSTRACT
The Repository of Assigned Search Tasks (RepAST) is a searchable repository created through a systematic review of the interactive information retrieval (IIR) research literature. It currently contains bibliographic details for approximately 750 articles, including empirical studies that employ assigned search tasks and a smaller number of conceptual papers on task-based searching. When available, the search task types, definitions and the task descriptions themselves are included. RepAST makes several contributions to the field. By bringing together examples of search task descriptions used in actual studies, RepAST provides a platform for studying practices within the research community and promoting greater conceptual clarity and consensus in the use of search tasks. To this end, the authors have published several studies based on analyses of the search tasks in the repository. In addition, researchers can use RepAST in a practical way, as a source of search task descriptions for reuse in new studies or in order to replicate prior research. In this interactive demo session, participants will have the opportunity to use the live RepAST system and provide feedback to the system designers.

Keywords
Assigned search tasks, research design, experimental user studies, interactive information retrieval.

INTRODUCTION
The use of assigned tasks as stimuli to prompt searching activities in studies of interactive information retrieval (IIR) has been common practice for decades. Such studies may be designed to examine the performance of the search system, the behavior of the searcher, or the effects of certain types of search tasks or task characteristics on search behaviors or outcomes. Search tasks can be defined as goal-directed activities carried out using search systems. Assigned search tasks normally take the form of a written description articulating a topic and the goal or purpose of the search task (Kelly, 2009). An important distinction has been made in the literature between search tasks, e.g., find the address of the nearest licensing office, and the work tasks in which they are embedded, e.g., obtain a driver’s license. Work tasks motivate and establish the parameters for search tasks (Toms et al., 2008; Wildemuth & Hughes, 2005).

Assigned search tasks play a central role in studies of IIR because they act as a proxy for the searcher’s information need and establish the goals and the criteria by which search results will be assessed. Examples of assigned search tasks range from specific fact-finding tasks, such as, “How much blood, on average, goes through the heart in one minute?” to broad exploratory tasks, such as, “Find studies that deal with the global warming and the reasons for it” (Aula, 2003). In some cases, participants are provided only with brief search topics and in others they are given detailed scenarios, including descriptions of the motivating work tasks, goals and expected outcomes (Borlund, 2003).

A common practice in assigning search tasks is to identify types or characteristics of search behaviors that are prevalent within a particular domain and to design the task descriptions to elicit these behaviors. Naturalistic studies are intended to observe people searching in their natural environments and completing real tasks, while experimental studies seek to isolate and observe particular effects on user behaviors. Because of this desire for control in experimental studies, researchers usually prefer to assign search tasks, which can be used to control the task effect by assigning the same tasks to all the subjects or to manipulate the task as an independent variable.

However, given the limited consensus and standardization in the design and implementation of assigned search tasks (Wildemuth & Freund, 2009), it is difficult to reach general conclusions as to the effect of search tasks on search behaviors or outcomes. Given that search tasks can vary along many dimensions, findings may be valid for a particular set of tasks, but we do not know to which additional tasks they may be validly applied. In order to make additional progress in experimental studies, we need...
to gain a better understanding of search tasks and their
effects. We need to be able to construct tasks having
particular attributes, knowing that our findings can then be
generalized to all search tasks having those attributes.

The Systematic Review of Assigned Search Tasks research
project (http://ils.unc.edu/searchtasks/index.html) was
established in response to these concerns, with the aim of
studying and documenting the use of assigned search tasks
in research studies. In this poster and demonstration, we
present an overview of the RepAST system and some of the
insights gained through analysis of the studies and search
tasks documented in it.

REPOSITORY DEVELOPMENT AND DESCRIPTION
For the past decade, a team of researchers has been
systematically scanning the literature to identify papers that
discuss or employ assigned search tasks. Topic-focused
literature searches have been conducted in Library &
Information Science Abstracts and the ACM Digital
Library. In addition, recent tables of contents of several
journals and conference proceedings were comprehensively
reviewed to identify potentially-relevant articles; these
included the Journal of the American Society for
Information Science & Technology (ASIST) (2010-present),
Information Processing & Management (2010-present),
Journal of Documentation (2010-present), Online
Information Review (2010-present), and proceedings of the
ASIST annual meetings (2010-present), ACM SIGIR
annual meetings (2011-present), the Human Computer
Interaction and Information Retrieval (HCIR) Symposium
(2007-present), Information Interactions in Context (IIiX)
biennial meetings (2008-present), and the Hawaii
International Conference on Systems Sciences (HICSS)
annual meetings (2012-present). Finally, as relevant articles
were identified, their reference lists were also examined for
additional potentially-relevant items.

RepAST now provides access to approximately 750
research and conceptual papers related to the search tasks
that are assigned in studies of IIR. It includes 577 empirical
studies of searching using assigned search tasks, 100
studies using naturalistic search tasks, and 62 conceptual
papers relating to search tasks. The distribution of papers by
publication year is presented Figure 1, showing that some
historical materials are included going back to 1969.
However, more than half the papers are from past decade,
reflecting the growth in task-based search studies in recent
years.

Each record in RepAST was created through manual
analysis of the full text of each article. In addition to the
citation of the paper, the data include the names and
definitions of the types of tasks assigned (quoted from the
original papers; no additional analysis of the tasks is
included in the repository itself. Data entry was carried out
by the Principle Investigators (PIs) and by graduate student
research assistants trained and supervised by the PIs.

![Figure 1. Number of Articles in RepAST by year](image)

The RepAST data structure consists of three files:

**Articles/Papers:** This file contains bibliographic
information (citation, abstract, and doi) for the articles
analyzed. In addition, it contains data on the type of article,
the domain and collection searched, the sample of study
participants, and whether outcome variables were analyzed
by task or task type.

**Task Types:** This file contains the names, definitions, and
other notes about the types of tasks assigned in the analyzed
studies, as specified by the author of the original paper. No
authority control over the task type names has been exerted.

**Tasks:** This file contains the full text of any of the assigned
tasks that are provided in the articles analyzed.

The RepAST search screen (Figure 2) offers a general
keyword search option as well as the ability to search
within specific fields, including the task types and
definitions, the text of task descriptions, the author names
and the citation. Searches can also be limited by the type of
article (conceptual/ theoretical, empirical-tasks assigned,
and empirical-naturalistic) and by whether or not the study
reports outcomes that are task dependent. The domain and
system/collection used in the study are also searchable.

An important feature of RepAST is the ability to search for
search task types and definitions and for search task
descriptions. The repository contains over 1200 task type
descriptions and many more search task texts. For
example, a search for “navigational” in the Task Type &
Definition field (see Figure 3) retrieves 11 papers, including
Broder’s (2002) paper that first defined navigational Web
search tasks. Among these results are 16 different examples
of navigational type search tasks that could be reused in
future studies.
APPLICATIONS AND NEXT STEPS

RepAST can be used to study the ways in which assigned search tasks are classified, designed and implemented. An examination of the list of search task types in RepAST shows a number of task types that are widely used, although the labels vary: complex vs. simple; general (subject, open-ended) vs. specific (fact-finding, known-item, close-ended); exploratory (browsing) vs. look-up; and transactional, navigational, and informational tasks (Wildemuth & Freund, 2009). These are the most common search task types; however, many others can be found in RepAST as well. These include system-specific or domain-specific search task types, compound task types that combine multiple task characteristics and others that serve the purposes of particular research studies. By providing a means of surveying implementations of search tasks in published research, RepAST offers guidance to researchers designing new studies as well as some empirical data to inform a growing body of conceptual work regarding search tasks.

Another application of RepAST is to search for particular search task types and compare their definitions across multiple studies. For example, as noted above, searching for navigational search tasks retrieves 11 papers and 16 examples of tasks. However, a closer examination reveals three distinct interpretations of this search task type: a) “the immediate intent is to reach a particular site” (Broder, 2002); b) “the searcher's purpose is to locate resources about a known topic” (Smith, Brian & Ashman, 2012); and c) search tasks associated with real world navigation needs, as in “You wish to go to the Espace Culture shop. You are a student at the University of Aix-en-Provence and you have no car. Which buses do you have to take?” (Chevalier & Kicka, 2006). Researchers wishing to employ “navigational” tasks should be aware of this variation and can use RepAST to identify and use the most widely accepted definition. From a methodological perspective, it is valuable to carry out in-depth analyses of particular task types, in order to resolve the definition of each type and allow researchers to use replicable task descriptions in future studies.

Finally, RepAST can be used as a library of reusable search tasks. Researchers can browse the available tasks and definitions and select some that are relevant for their own work. While reuse of search tasks is relatively rare in information science, it has the potential to strengthen the external validity of research findings through comparison of results across systems or study populations. Furthermore, the design of assigned search tasks can be quite challenging, especially when the intent is to manipulate particular task characteristics and to induce naturalistic searching behaviors. The opportunity to share tasks through
RepAST may encourage researchers to invest in designing better tasks themselves and to make greater use of carefully-designed existing tasks.

CONCLUSION
The long-term objective of creating the Repository of Assigned Search Tasks (RepAST) was to assist IIR researchers in improving the design of the task descriptions they implement in IIR experiments. We hope that the Repository will have two types of impact: 1) IIR researchers will be able to use RepAST to identify task descriptions that fit their needs and can be re-used in new studies; and 2) methodological studies can focus on particular task attributes and task types, allowing researchers to focus their attention on those task attributes or types that have the greatest impact on search success.

There are some limitations in the current implementation of RepAST of which researchers should be aware. While the collection is extensive and provides good coverage of the information science and information retrieval literature, it is not comprehensive, as that would require a much broader and deeper survey across multiple cognate disciplines. A second limitation stems from the arm’s length approach taken in collecting this data. Rather than assessing or exerting control over the use of search tasks in these studies, the task descriptions and definitions are presented as they were originally published. Therefore, it is up to the researchers using RepAST to determine how best to interpret and make use of this data.

We look forward to presenting this resource to the community and hope to receive input on its design and potential usage. We hope that RepAST can play a role in building our knowledge of people’s searching behaviors and can support the design of IR systems that enable searchers to be successful.

ACKNOWLEDGMENTS
RepAST could not exist without the contributions of many graduate research assistants, including: Ariel Deardorff, Chris Doty, Ying Han, Amanda Leinberger and Jung Sun Oh.

REFERENCES


