

Bringing search into task: Identifying task primitives

Fernando Diaz, Catherine Smith, Simone Stumpf, and Elaine Toms

Problem / Motivation / Goals

In work on interactive information retrieval, search is often studied in isolation from the task activities that prompt searching and which are dependent on the products of searching. *In situ* log studies describe search behavior that occurs in natural *task contexts* (e.g. planning a wedding), however these studies often lack information about that context, as well as data on the task activities that occur outside of the search system (e.g. recording information in a word processor). In experimental studies, research subjects are generally given a task context (e.g. pretend you are a journalist) and assigned specific task goals (e.g. find background information on the sequester). In most studies in this setting, task activities that might precede and follow search are hypothetical and little is learned about the dependencies between these activities and search. Longitudinal studies conducted in specialized domain settings (e.g. chemical engineering) have examined *in situ* task and search activities however, few studies have attempted to systematically extend this work across domains. In summary, little is known about the fundamental characteristics of dependencies between task activities and searching (hereafter called *task dependencies*). This knowledge is essential to the development of the task models required for support integrated search and task activities.

The investigation of task dependencies across domains requires a standardized framework for decomposing task structure to a sufficient level of detail. Because a suitable framework does not yet exist, part of the problem is to develop and test an initial framework. In applying the framework, the goal is to discover and describe task dependencies that occur universally across domains and task contexts; we term this type of dependency a *task-primitive*.

We propose to examine the question of whether there exists a set of task-primitives and to address the following research questions specifically.

Research questions

- Where in the flow of sub-task activities is search invoked?
- Can a set of task primitives be identified across diverse contexts and task activities?
- Can common dependencies between primitives be identified?
- How can descriptions of task primitives be standardized for research purposes?
- What might be the components of a generalized task activity model?

Approach

Our approach is to record, decompose, and analyze *in situ* task activities at a level of specificity sufficient to reveal task primitives. We will study two diverse task contexts, and will decompose task structure using the framework summarized below.

Data collection overview

Because we seek to understand naturally occurring task dependencies, a longitudinal design is required. Data will be collected using multiple methods and from multiple sources, including initial and periodic interviews about the task under study, collection of comprehensive log data from multiple devices, and participant diaries and annotations. Interviews will focus on expected and experienced task structure. Log data will record transitions between search systems and systems used in completing the task, as well as selected transactions within select systems. Periodically

throughout data collection, participants will be asked to record annotations on specific aspects of task activities. Interviews, diaries, and annotations will be transcribed and coded. Coded data will be integrated with log data for analysis.

Selecting task contexts

We define a *task context* as a high-level multi-part goal. Task contexts differ in many dimensions. For the proposed study we focus on the specificity of the goal, level of collaboration, time urgency, duration, and geographic scale. Examples of task contexts include crisis response, travel planning, buying a house, selecting a school, developing a product, debugging code, managing a medical condition, handling a family legal matter, etc..

For the proposed study we have selected two task contexts that vary on several dimensions. The first, a small group assignment to be completed in an online course, involves collaboration with time urgency and limited duration. The second task context, personal management of an asthma condition, is primarily individual with a regular ongoing repeating time urgency. Both have specific articulated goals and limited geographic scale. The differences between the contexts provides the contrast required to address the question of whether task-primitives can be found in diverse contexts.

The framework and its application

In order to standardize the decomposition of task structure, we will use a three-phase iterative analytical framework. The framework sets aside questions of task detection or classification of tasks by type. Below, we step briefly through the application of the framework for analysis.

Phase 1: Task context decomposition

The table below outlines the two task contexts and examples of possible associated *task activities*, which we define as high-level multi-part sub-tasks of the task context. As defined in this initial version of the framework (we expect that the framework will evolve as we learn about task dependencies), task activities express the steps or composition of the larger context. For the asthma study the activities repeat regularly, while for the group project they have a finite time limit. It is expected that task activities will emerge during iterative application of the framework.

| <i>Phase 1: Task activity decomposition</i> | TASK CONTEXT | |
|---|---|---|
| | Group Project in Online Graduate Course | Asthma Management |
| TASK ACTIVITIES (hypothetical) | <ul style="list-style-type: none"> • Select topic from list • Plan project • Assign duties • Conduct research • Share and analyze • Prepare paper | <ul style="list-style-type: none"> • Measure peak flow rate • Observe symptoms • Observe triggers • Adjust medication |

Phase 2: Task activity analysis

The table below provides an example analysis of integrated data from one hypothetical task activity in the group project task context. The goal of analysis is to identify activity sequences that comprise task activities, and with those activities, to identify potential task primitives. It is expected that primitives will emerge during analysis of the integrated data. Analysis will be conducted using a combination of algorithmic and human processes.

In the example below participant Joe has used his laptop to search for information on the assigned project topic, looking a library catalog and a database service. He’s recorded and saved notes using his word processor. Four days later Joe uses his cellphone to search the web, and browses a Wikipedia page and a YouTube video, all related to the project topic. The next day he sends an instant message to Jill, his partner on the project and she replies. She then uses her laptop to search a database service and saves PDFs of several articles to a remote server. As can be seen in the rightmost column, each sequence has been identified as a potential task primitive.

| <i>Phase 2: TASK ACTIVITY ANALYSIS (example task activity: Select Topic)</i> | | | | | | Pot. TASK PRIMITIVE |
|--|--------------------|---------------|------------------|--------------------------|---|---------------------|
| <i>time</i> | <i>participant</i> | <i>device</i> | <i>tool used</i> | <i>information input</i> | <i>information output</i> | |
| AM 9/8 | Joe | laptop | Library catalog | topic descriptions | – bibliographic records | explore |
| | | laptop | Database service | topic descriptions | – bibliographic records – article abstracts – full text document displays | |
| | | laptop | Word processor | notes on topics | – text file | |
| PM 9/12 | Joe | phone | Search engine | topic terms | – results pages | |
| | | | Wikipedia | links | – text pages | |
| | | | YouTube | topic terms | – video | |
| 9/13 AM | Joe | phone | message | comment on progress | – message to Jill | contact |
| 9/13 AM | Jill | phone | message | message from Joe | – comment on plan | |
| | | | message | comment on plan | – message to Joe | |
| | | laptop | Database service | topic descriptions | – bibliographic records – article abstracts | gather |
| | | laptop | remote server | full text documents | – saved PDF files | store |

Phase 3. Analysis across task contexts and activities

In the third phase of analysis, potential task primitives will be aligned across task contexts and task activities, with the objective of identifying and describing any global characteristics for each primitive, and any features that may be task dependent.

| <i>Phase 3</i> | TASK ACTIVITIES | | | | | |
|---------------------------|-----------------|-----------|---------------|------------------------|-----------|-------------------|
| | Group Project | | | Asthma Management | | |
| Potential TASK PRIMITIVES | select topic | ...etc... | prepare paper | measure peak flow rate | ...etc... | adjust medication |
| explore | | | | | | |
| gather | | | | | | |
| store | | | | | | |

| | | | | | | |
|-------------|--|--|--|--|--|--|
| route | | | | | | |
| filter | | | | | | |
| evaluate | | | | | | |
| share | | | | | | |
| compare | | | | | | |
| integrate | | | | | | |
| extract | | | | | | |
| separate | | | | | | |
| find trends | | | | | | |
| contact | | | | | | |
| | | | | | | |

Challenges / Resources Required / Caveats

In order to capture task activities on all types of devices, log recording methods are needed. If these methods are not available, knowledge of task activities will be limited to those that occur on personal computers. A second challenge is participant recruiting and retention for long-term ubiquitous data collection. Because participation requires the disclosure of information likely to be perceived as personal and private, adequate incentives will be necessary. Also, because the data to be collected has a large scope and will be voluminous, integration and analysis will be complex and will require resources.

Plan for future

The proposed study is limited to two task contexts in two different domains. Task contexts vary on many other dimensions. Assuming the analytical framework proves to be a useful approach, further research in different domains and contexts will be needed to validate and extend the task primitives found in the first study.

Impact

Interactive search is ubiquitous in everyday life. It is integral to and useful for all types of life goals. At present, search is not well integrated with the tasks it supports. This makes information intensive tasks burdensome, particularly for populations that lack experience with the many applications often required to support task completion. By identifying and describing task primitives we may be able to determine those which have the broadest application and thus prioritize research and development efforts.