Search Situations and Transitions

Nils Pharo
Oslo University College, Norway

Kalervo Järvelin
University of Tampere, Finland

INTRODUCTION

In order to understand the nature of Web information search processes it is necessary to identify the interplay of factors at the micro-level, that is, to understand how search process related factors such as the actions performed by the searcher on the system are influenced by various factors that might explain it, for example, those related to his work task, search task, knowledge about the work task or searching and etcetera. The Search Situation Transition (SST) method schema provides a framework for such analysis.

BACKGROUND

Studies of information seeking and information retrieval (IS&R) have identified many factors that influence the selection and use of sources for information seeking and retrieval. What has been lacking is knowledge on whether, and how, these factors influence the actual search performance. Web information searching often seems to be a rather haphazard behaviour where searchers seem to behave irrationally, that is, they do not follow optimal textbook prescriptions (e.g., Ackermann & Hartman, 2003).

In the research literature it is claimed that factors related to the searcher's personal characteristics, search task, and social/organisational environment influence the searcher during his selection and use of information sources. These factors have been classified and discussed in great detail in the literature, and more recently the searcher's work task has been focused on as playing a major role (e.g., Byström & Järvelin, 1995; Vakkari, 2001). The SST method schema focuses specifically on the search process and how it is affected by external factors.

There are several studies that focus on search processes in other information systems (e.g., Marchionini et al., 1993). These studies have primarily been based on logged and/or video taped data of online bibliographic searches. However, their scope has been search tasks and searcher characteristics, focusing on term selections and results evaluation. Similar examples can be found in the Web searching context (e.g., Wang & Tenopir, 1998; Fidel et al., 1999; Silverstein et al., 1999; Jansen, Spink, & Saracevic, 2000); these studies analyse characteristics of the Web Information Search (WIS) processes, such as term selection, search task strategies and searcher characteristics, but do not aim at explaining the process itself and the factors that guide it. The current method schema focuses on explaining the process at a micro-level.

Early studies of Web searching have to a large degree used log analysis (see review in Jansen & Pooch [2001]) or surveys (e.g., GVU's WWW user surveys [2001]) as their data collection methods. Log analysis can provide researchers with data on large numbers of user-system interactions focusing on users' actions. Most often log analysis has been used to see how searchers formulate and reformulate queries (e.g., Spink et al., 2001). The user surveys have focused on demographics of Web users and collected information on the use of different kinds of Web resources, time spent on Web use, e-shopping, etcetera. Although both these kinds of methods may reveal important information about how and why people use the Web, they are unable to point out what causes the searcher to perform the actions he does. We cannot use these methods if we want to learn how work tasks, search tasks, and searcher's personality directly affect Web information search processes. The SST method schema (Pharo, 2002; Pharo & Järvelin, 2004) was developed for such analyses.

MAIN THRUST

To present a method (e.g., Bunge, 1967), as well as a method schema (Eloranta, 1979), one needs to define its domain, procedure and justifications (Newell, 1969; Pharo, 2002). Both the domain and procedure will be presented below in order to clarify the usability of the SST method schema.

The Method Schema's Domain

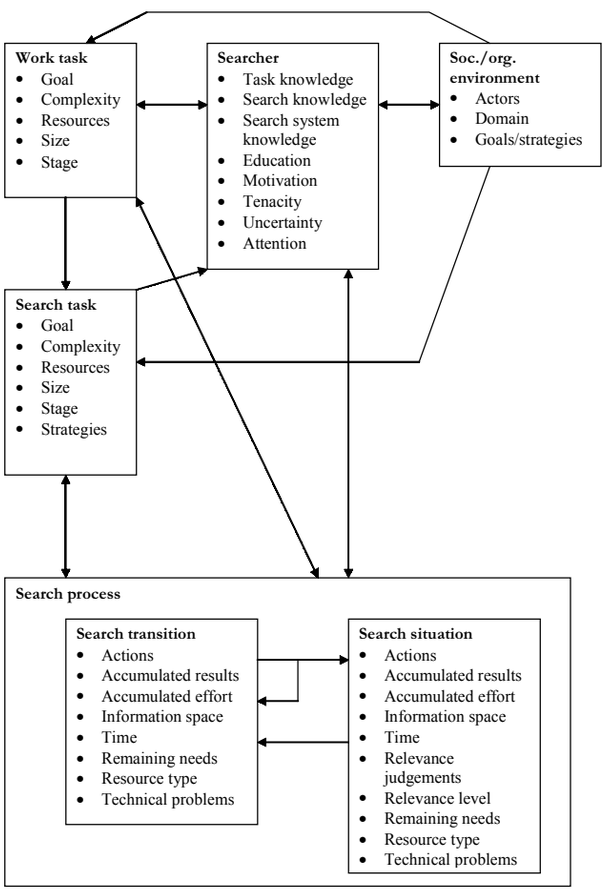
The problem statement, or *domain,* which is used in the following, states the properties of the problem the method is intended for and their relationships. This designates how general it is possible to make the procedure for handling the problem.

Figure 1 is a representation of the framework's five *categories* and the relationships existing between them.

The *search process* category consists of two subcategories: *search situation* and *search transition*. The search process category will be emphasised here, the other categories and their attributes are well known from the IS&R literature (for details see Pharo, 2002)

Search *situations* are the periods during a search process when the searcher examines a resource in order to find information that may be of help in executing his work task. Situations may take place in the same kind of resources as transitions depending on the search task; if the searcher wants to learn more about the structuring of subject indices it would be natural to examine such resource types for that purpose.

Figure 1. The conceptual framework - the domain of the method schema



Search *transitions* are executed in order to find resources in which the searcher believes there may be information that can help execute his task. The transitions consist of source selection and inter-source navigation. An alternative way of explaining this is to say that while situations represent interaction with real information the transitions deal with *meta*-information.

Action is used to describe the moves made by the searcher during a situation/transition. In Web interaction this includes the following of links, entering of queries, and reading of pages. The actions may be influenced, for example, by a search task strategy.

The *accumulated results* refer to the information already found. This includes information found in previous situations as well as information found in the current one. Accumulated results relate to the completion of the information need (or the futility of trying this).

The *accumulated efforts* refer to how much work the searcher has had to invest from the start of the present session (or in prior sessions) up to the current position. In addition it can refer specifically to effort invested in the current situation.

The *information space* refers to the part of the Web that the searcher has navigated, as well as the information space anticipated by the searcher. The searcher has developed a cognitive model of the information space based on his knowledge about the Web and the existing resources on the Web, but also on his knowledge about institutions and organisations that he expects to be represented on the Web.

Time can be used to specify how the total amount of time spent during a search process influences the current situation, but it can also relate to the specific time used in that situation.

The *remaining needs* refer to what the searcher has planned to search for in the continuation of the search process and possibly in subsequent search processes.

Web *resource types* differ from each other with respect to content and format. Some are known from the world of paper-based publishing, such as newspapers, scientific journals, dissertations, novels, and collections of poems, but there are many new genres that have originated on the Web (home pages, various kinds of interactive resources, etc.) (Shepherd & Watters, 1998).

"Technical problems" is used to describe problems caused by the software in use, both on the client and server sides of interaction. Lack of bandwidth may also cause problems, for example in accessing resources that heavily depend on transmission of large amounts of data. Web pages that have disappeared also cause this kind of problem.