

The Information Seeking and Retrieval process at the Swedish Patent- and Registration Office

Moving from Lab-based to real life work-task environment

Preben Hansen

SICS – Swedish Institute of Computer Science and
Tampere University, Dept. of Information Studies, Finland

Kalervo Järvelin

Tampere University, Dept. of Information Studies, Finland

preben@sics.se and likaja@uta.fi

Abstract

The following paper describes a set of methods that is currently used in a study of the task performance process of patent engineers within the Swedish Patent- and Registration Office (SPRO)¹. The focus of the study is to investigate the relationship between the user's work-task and the information seeking and retrieval process. The study is performed within a real life work setting where patent engineers are performing real work tasks involving real information needs. This paper will focus on and describe a set of data collection methods used in our study. Generally, IR studies are performed within a controlled laboratory environment with controlled variables and design or simulated information need. We argue that we need to take a broader perspective on the information seeking and retrieval in order to understand the task performance process and elicit requirements for information systems design.

Introduction

Research on traditional IR techniques and tools cannot solely provide the understanding and knowledge of the interaction between the user and the IR system as pointed out by Belkin et al. (1995), Ingwersen (1992, 1996) and Saracevic (1996). It is now rather well understood that *information seeking* in electronic environments (Marchionini, 1995) is a dynamic process driven by an individual's or a groups need for information.

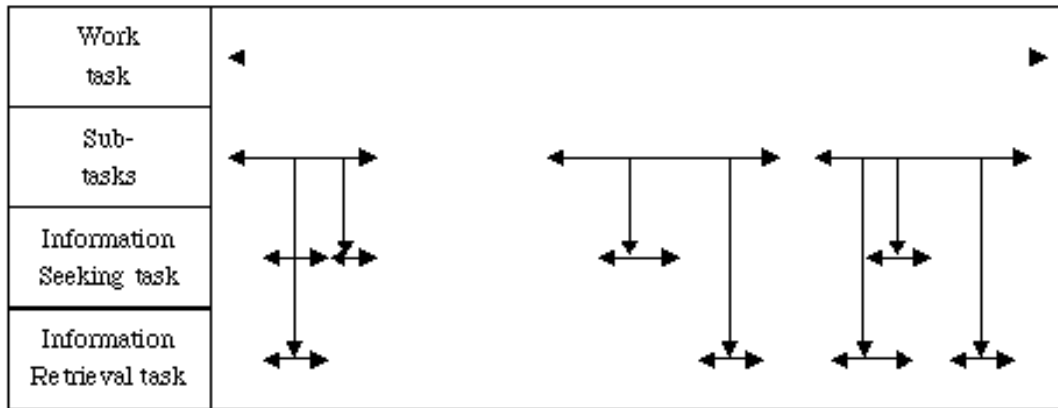
The information seeking and retrieval research has focused on issues such as information

seeking strategies (e.g. Bates, 1989, 1990; Belkin et al. 1993, 1995) and user behaviour (e.g. Borgman, 1989; Kuhlthau, 1993a, 1993b, Wilson, 1997), as well as issues on design of IR systems and user interfaces (e.g. Rouse and Rouse, 1984; Bates 1989; Belkin et al. 1993; Brajnik et. al, 1996). Furthermore, it has been argued that users' work tasks and goals (Rasmussen et. al., 1994; Byström and Järvelin, 1995) also need to be taken into account and understood in the framework of information seeking and retrieval.

Our point of departure is that the *information retrieval* activity is *embedded* in the *information seeking* process. Furthermore, within a professional setting such as the Patent domain, the information seeking and retrieval activities are embedded in *work-tasks* [Figure 1]. In order to acquire a wider understanding of the processes that undergo IR, we need to adopt a broader and contextual perspective, which includes the tasks people are performing. Furthermore, there are also (not treated in this paper) the *system designed task* and also the *interaction tasks*.

¹ A security agreement supports all activities with the SPRO.

Figure 1. Task dimensions



In order to elicit requirements for the design of information systems, access tools and user interfaces, we need to understand the task performance process involved and the user's tasks characteristics within an information seeking and retrieval environment. To do this, we need to take a broader viewpoint, not only the analysis and evaluation of the query-document match (system performance), but also include contextual and situational factors such as the user and her knowledge levels and information need; work-tasks and their characteristics and types, information sources and information types; and information seeking and retrieval success and use of retrieved information.

We need to acquire knowledge about the tasks that people do or want to do and therefore we need to understand the task performance process. Support should be designed to provide the user with the necessary assistance in reaching her goal when performing her task. Some general problems addressed within the current study are:

- The *concept of task* is central to user-centred system design within the HCI, but

is seldom taken into account for the design of IR systems and user interface.

- The *context of tasks* is vague and not well understood and therefore many IR systems have problems supporting both tasks and users
- There are relatively few studies that *empirically* connect the work task to the information seeking and retrieval process.
- Information seeking and retrieval should not be treated in isolation, rather as *embedded* in a task context.

Research goals

In short, our general goal is to investigate the relationship between work-task situations and users information retrieval processes in order to suggest guidelines and requirements, which may be used for the design of interactive IR systems.

- how work-tasks related to the information seeking and retrieval process?
- how do types of tasks (patent applications) influence the information seeking and retrieval process?
- how may knowledge about the task and task characteristics influence the design of IR tools and systems?

Information Seeking and Retrieval – A work task perspective

Various types of IR systems and IR techniques are applied to support different search strategies serving different interaction modes in the user interface design. *Information* (or information objects) has different representations and structures. There are different information *sources* (collections of objects) and users generally interact with information at different *levels* (full-text or bibliographic information) and in different *media* (text, and image). Users interact with information as single objects or as groups of objects. *Users* have different preferences, experience and knowledge of the subject domain and IR. Users also apply different approaches and information seeking strategies when performing an information-seeking task (Belkin et. al., 1993). Furthermore, research has pointed out that we also need to consider the *user's workspace*, containing the user's *work task*, and that these are related to the formation of information needs (Rasmussen et. al., 1994). The different goals and characteristics of the task may influence the way the user approaches the information seeking and retrieval activity. Thus, all these components are important factors affecting the task performance and the information seeking processes.

Task-based approach

Bennett (1972, p. 189) concludes that the challenge is to transform theories into engineering by developing an “agreement on ways for characterising user tasks, for allocating interface resources to meet task requirements and for evaluating user effectiveness in task performance”. It is obvious that different task characteristics will influence the information seeking and task performance process in various ways. Task characteristics have been discussed in different contexts such as: Simple - Complex tasks (Byström and Järvelin, 1995); Structured - Unstructured tasks (O'Day and Jeffries, 1993); User - Computer controlled (O'Day and Jeffries, 1993 and Bates, 1990); Active - Passive tasks; Routine tasks (Hill et al., 1993); Single – Multiple tasks (Preece, 1992 and Hill et al., 1993, Belkin et al., 1993, Smith et al., 1997); Task continuity - Discontinuity (O'Day and Jeffries, 1993); Task uncertainty (Kuhlthau, 1991); Perceived tasks (subjective and objective level; Byström, 1999); Defined – Muddled work task (Ingwersen, 1996, 1997).

See Hansen (1999) for a more detailed discussion.

Tasks are a fundamental concept for information science as well as for Human-Computer Interaction (HCI) and system design. An IR system may provide good performance for one information seeking strategy for solving one specific information need and task, but provide poor support for other strategies and tasks. We need to understand the user's tasks and the task context in relation to the users information seeking and retrieval activities.

Little empirical work has been done in this area, although it is obvious from the findings above that different tasks influence different needs and constraints on the human to computer interaction. What is not so well understood is how to meet these requirements. We need more *descriptive* models of tasks in real life situations that explain the relationship between different tasks and information seeking in relation to IR activities. Such an analysis is particularly important to *inform* in the design process of information access systems.

The task performance process

The task performance process may have both **contextual** and **situational** aspects (Hackos, 1998). In professional life, such as the Patent work environment, a context may be defined as the environment or domain in which different work task situations exist. Within these different work task situations may different information seeking and retrieval activities be embedded.

- *Contextual attributes* may be types of tasks, domains, work-task situations, processes, actors and the organisational factor such as constraints.
- *Situational attributes* may be knowledge levels, types of information needs, information sources, types of media, information seeking strategies, IR techniques, level of success, time, information use (e.g. Ingwersen 1992 and Preece et. al., 1994).

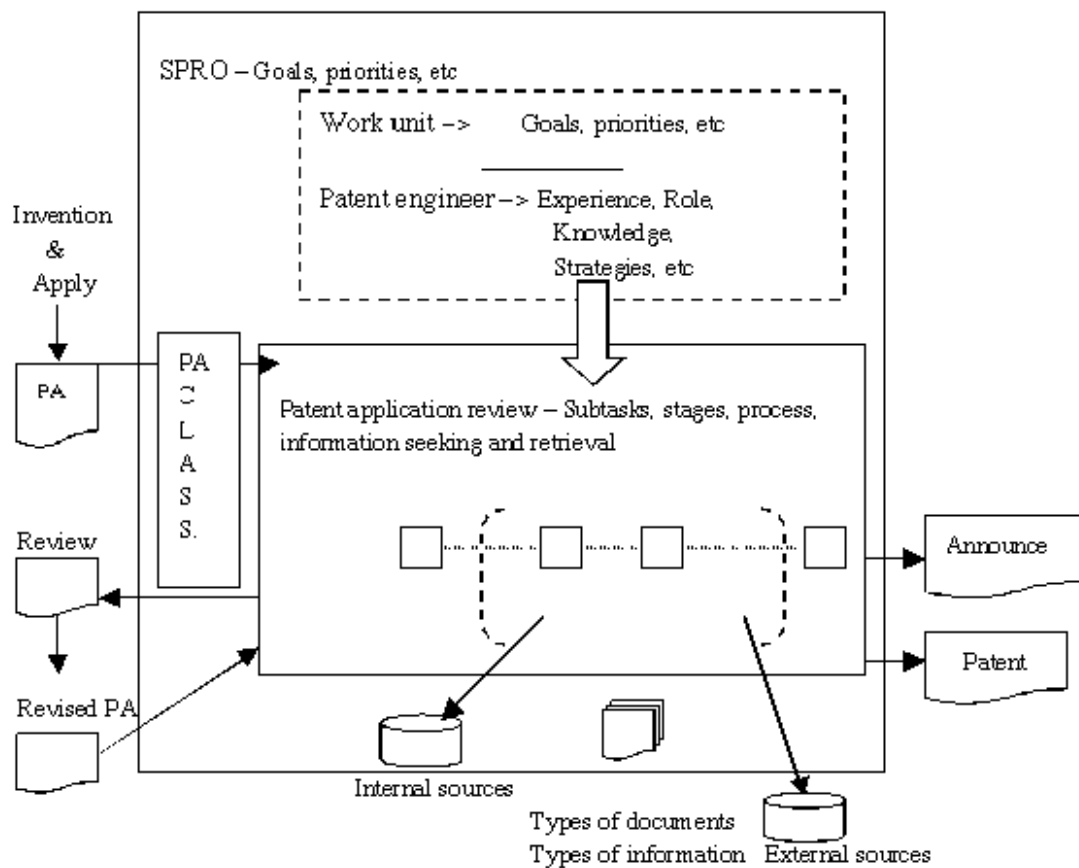
A task (or sub-task) performance process has, more or less, an identifiable or recognizable starting point and an ending point. From a contextual point, the work-tasks that need to be fulfilled within a workplace may be guided by organisational factors such as guidelines, restriction and legal aspects. The work-task may be set, externally or internally, by a person, a group of persons, or by an

organisation. From within the workplace, there may exist a predefined set of work-tasks that needs to be performed. The work-tasks, in our case the patent applications, are then classified and forwarded to the appropriate group of patent engineers. Finally, the patent application task is assigned to the task performer, that is the patent engineer that will handle the patent application (Figure 2).

At the situational level, the work-task performer is faced with a specific *work task* in a specific situation. The work-task to be performed may have several sub-tasks, as in the case of SPRO. The work-task may involve different information needs; knowledge levels

that may lead to information seeking activities which involve different information sources and information seeking strategies. The task performance process may or may not involve the use of information retrieval systems. Often, in the case of patent engineers, an IR task process is initiated as part of the problem solving activity. Finally, the task performance is stopped based on the feedback and assessment of the retrieved information in relation to the perceived information need of the work-task.

Figure 2. General framework of the handling Patent application task.



Methodological approach

Our design approach in this on-going study is based on the following recognized shortcomings within IR research. There are only a:

- few studies on people performing real life information seeking tasks with real information needs, and
- few studies are done in a real-world online IR setting which seeks to analyze and develop IR systems from the task and seeking perspective

We then made the following decisions regarding our present study:

- to implement the study in real work situations with users performing real information seeking tasks with real information needs
- to involve actual and multiple IR and other access systems.
- to identify and describe the task performance process of patent engineers through monitoring and observing their activities on-site.
- to collect both qualitative (interviews electronic diaries and observations) and quantitative (log-statistics) data from users performing information seeking and retrieval tasks.

The project is a collaboration between the Swedish Patent- and Registration Office (SPRO) and SICS. Initially, we wanted to perform our study in a setting or domain that were “IR-intensive”, that is, an environment that contained a relatively large amount of information seeking and retrieval activities. We found that the SPRO performed information seeking and retrieval activities on almost a daily basis. Another factor that attracted us was that they were using different types information sources. Furthermore, since we wanted to study the task process of solving patent applications, the frequency of IR activities was important. The study was designed to collect data from real-life work situations, which involves IR-intensive activities. The study intends to be both open and holistic in its approach.

10 professional patent engineers from SPRO participated in the study. The informants were selected by the SPRO. The selection was guided by two general criteria: representation of different technology areas and different years of working experience at SPRO. Since we were interested in task performance process and especially the relationship of work-tasks and the information seeking and retrieval process, we needed to use methods that made it possible to monitor the work-task processes. Below, we present the different methods used in the study:

Group introduction

Information about the components of the project. The group introduction served two purposes: a) the opportunity for the researcher to present the study and its goals, and b) the opportunity for the participants to identify the

others involved in the study and to ask any questions concerning the study.

Pre-Interview session

“Theme-based” and open-ended interviews. A set of question were designed to collect data about demographics, data on experience and knowledge levels, contextual factors and descriptions on how they usually search for information, what sources they use, how they use information, etc. The interview was performed in an informal non-hierarchical way. The questions pre-defined served as “bookmarks” upon which the interview was based on

Data collection: written notes and tape-recording.

Electronic “Diary”

The diary was constructed containing a set of suggested stages/steps in a proposed task process (e.g. there were questions about the starting phase and the ending phase). These stages/steps are presented as suggestions and the goal is to collect data about the construction, performance and ending of each sub-task. It also contained a set of reminding keywords to check during the note taking time by the participant. The diary also contained an empty field for logging information on online sessions. The Diary was set up electronically. The participants were asked to send in their diaries at the end of each workday.

Data collection: written statements and descriptions of process and search logs from different information systems during the whole task performance process.

Participatory and Focus observations

In parallel, ongoing and continuous on site visits are done, observing patent engineers in their work with patent applications. The researcher used a list of key questions to be asked when appropriate. Awareness of unexpected situations and activities emphasized. Whenever a subject made a move relevant for the study, the researcher stepped into the scene and asked about that “shift” in the task performance. The subjects were allowed to “talk aloud”. This means that in parallel to the participants writing of “Diaries”, the researcher moved around, in an “ethnographical” way, at the SPRO every day observing people performing their tasks.

Data collection: written notes.

Post-interview

If necessary, follow-up interviews will be done to clarify any specific issues.

Data collection: Discussing the notes for additional information about the outcome

The data collection period was 5 weeks (May to late June 2000). The problem with the patent application, from our point of view, is related to time. It takes up to 2 years to complete one single patent application. This would not be feasible within our study. Instead, we decided to focus on sub-tasks within different patent applications. These had a "life-time" of approx. 1-4 days. We decided that it would be a suitable work unit to be observed. The preliminary goal is to collect the description of approx. 50 patent application tasks from the electronic diaries and to physically observe around 10-15 sub-tasks. These tasks will be followed and observed in full as much as possible. Prior to the main study phase, a pilot study was conducted in order to validate the methods. After the pilot test, some changes were made regarding the "Diary" and the keyword-scheme used during the observation. We will perform both qualitative and quantitative analysis on the data collected.

Categories of variables

The following different categories are seen as needed as relevant to the task process in the present study. Each category contains several different variables that will be used for the analysis of qualitative and quantitative data in the study. The categories are divided into either dependent or independent variables:

- *Domain / Context-related variables* (Constant)
- *Actor* (Patent engineer) (Constant)
- *Task* (Variable)
- *Information Need* (variable)
- *Information Source* (Variable)
- *Type of Information/Media/Document* (Variable)
- *Information Seeking Strategies* (Variable)
- *IR techniques* (Variable)
- *Success* (Variable)
- *Information Use* or Use of information (Variable)

Preliminary results

Since the period of data collection is still in its final stage, we cannot communicate any reliable results based on the data so far. The data need to be analyzed and interpreted and thus it is not possible to say anything concrete. However, some indicative observations may be presented from the study at SPRO.

Information Retrieval is generally understood as an individual-driven activity. In this paper, we would like to highlight the aspects of collaborative IR. This area was not the focus in our study, but nevertheless, it showed some interesting finding regarding the task performance process and IR activities. These activities were mainly captured through the method of observation. Our first preliminary findings points to that the patent engineers are involved in different kinds of collaborative IR activities. The activities may be categorised along the following dimensions:

- collaborative IR that are related to internal or external activities
- collaborative IR that are related to individual or group related activities

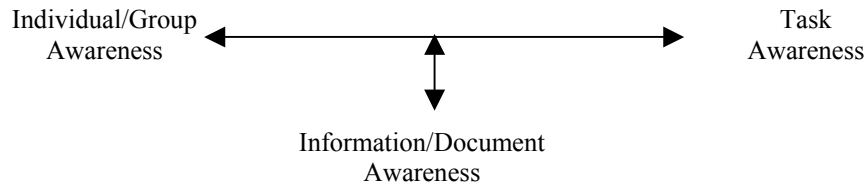
With internal we mainly mean activities within the national patent office and external then between national patent offices and between the national patent office and other national bodies. Examples of internal activities may be activities such as the re-use of information seeking strategies and classification codes used in searching for a specific class of patent applications. This may be done individually or within the work group/unit.

International patent applications handled in one country and then forwarded to a second country are example of external collaborative IR activities. The references and pointers to documents made by the first country are re-examined and re-used in retrieval activities in the second country.

Patent applications may be a work done by one person or by several. Due to the long period of handling of a patent application, this may lead to that different sub-tasks are divided between two persons even though this is not recommended. Sometimes, due to time pressure, a patent engineer has to work with an application not within her/his subject area. This may lead to a higher degree of collaborative activities with the group to which the patent application belongs.

On the document/information level, terms, classification codes and concepts from previous or in parallel performed IR activities may be either re-used or act as guidance in a current task.

Based on the preliminary findings, we may construct a preliminary and general model that illustrate the major interacting activities involved:



Finally, it is important to mention that these are just a few examples of collaborative IR activities and that more extensive analysis certainly is needed before we can say anything more specific. However, these preliminary findings points towards aspects that might be considered in the design of IR systems. In order to support collaborative activities within the information seeking and retrieval, we need to design tools that will support users and groups of users when interacting with information access systems.

Discussion

The discussion mainly concerns methodological issues, since it is to early to draw any conclusions on the data at this stage of the study, although we made some preliminary notes in the previous section. The study is currently in the end of the data collection phase and hence, should be regarded as work in progress. The analysis of data will be done during the autumn 2000.

The discussion in this section will mainly concern methodological issues, since it is to early to draw any conclusions on the data at this stage of the study, although we made some preliminary notes in the previous section.

At this phase of the project, some issues might be put forward concerning the methods used in the study:

- Generally, the methods used in the study created a large amount of logistics e.g. rescheduling dates for visits and other communications with the participants.
- The methods used made it possible to move the study from a controlled lab setting to the real work setting and to collect empirical data on the work tasks of patent engineers at SPRO.
- The combination of online diaries and focus observations made it possible to manage a complex data collection situation alone and within a reasonable timeframe.

- During the study, the patent engineers performed their work in a natural way without any major interruptions.
- The focus observations created a very rich "map" of activities and aspects of the work of the patent engineers.
- The chosen combination of methods made it possible to "enter" and "step out" of the work environment rather simple. When the planned set of cases were collected, we just send a message to all the participants saying that they now worked on their last case to be reported and after that, they did not need to submit anything more.
- The data collected contains written notes by the researcher, tape-recorded interviews, log statistics from online IR sessions, and written notes during focus observations. All these different types of data now need to be transcribed and analysed. Together they will provide a rich and complex set of data. This will probably be time consuming.
- The proposed methods made it possible to monitor how the patent engineers treated the patent applications and how the information need/problem were formulated. Furthermore, the methods made it possible to observe the IR strategies that were chosen and how the patent engineers made decisions concerning the specific application.
- The methods proposed, especially the observations uncovered a detailed picture of the subject's information seeking strategies as well as their personal thinking in specific situations.
- The combination of the "diary" and observations methods together made it possible to gather information about a large set of patent retrieval tasks. The diaries contained a predefined set of "keywords" that "guided" the diaries, and thus gathered similar data sets across all 10 subjects. The information gathered through observations were more individual and made it possible to collect data on specific activities that may be used together with the diaries in the later stage of analysis.

- Furthermore, together with the "diaries" logging information were gathered. The methods of "diaries", observations, and logging, created three different data corpuses to be treated and analyzed. Another task in our study will thus be to integrate these three kinds of data sets.

Acknowledgement

I would like to thank the Swedish Patent- and Registration Office (SPRO), especially C-O Gustafsson and Stig Edhborg, and all who participated in the study for their cooperativeness. I also want to thank Jussi Karlgren and Annika Waern at SICS for their support. We also want to thank the anonymous reviewers for their valuable comments on the first draft of the paper.

References

- Bates, M. (1989). The design of browsing and berrypicking techniques for online search interface. *Online Review*, 1989, Vol. 13, No. 5, pp. 407-424.
- Bates, M. (1990). Where should the person stop and the information search interface start? *Information Processing & Management*, 1990, Vol. 26, No. 5, pp. 575-591.
- Belkin, N. J., C. Cool, A. Stein and U. Thiel (1995). Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems With Applications*, 1995, Vol. 9 (3), pp. 379-396.
- Belkin, N. J., P. G. Marchetti and C. Cool (1993). BRAQUE: Design of an interface to support user interaction in Information Retrieval. *Information Processing & Management* Vol. 29, No. 3, pp. 325-344, 1993.
- Bennett, J. L. (1972). The user interface in interactive systems. *ARIST*, 7, pp. 159-196.
- Byström, K. and K. Järvelin (1995). Task complexity affects information seeking and use. *Information Processing & Management*, Vol. 31, No. 2, pp. 191-213.
- Byström, K. (1999). Task complexity, information types and information sources. Tampere, Finland: University of Tampere, Acta Universitatis Tampereensis 688.
- Borgman, C. (1989). All users of information retrieval systems are not created equal: an exploration into individual differences. *Information Processing & Management*, Vol. 25 (3), pp. 237-252.
- Brajnik, G., S. Mizzaro and C. Tasso (1996). Evaluating user interfaces to information retrieval systems. A case study on user support. In: Frei, H-P., Harman, D., Schäuble, P., and Wilkinson, R. (eds.). *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'96)*. Zurich, Switzerland. August 18-22, pp. 128-136. 1996.
- Hackos, J. and J. Redish (1998). *User and Task Analysis for Interface Design*. New York: Wiley.
- Hansen, Preben (1999). User Interface design for IR Interaction. A Task-oriented approach. COLIS 3, Dubrovnik, 1999.
- Hill, B., J. Long, W. Smith and A. Whitefield (1993). Planning for Multiple Task Work. An Analysis of a Medical Reception Worksystem. In: Ahslund, S., et al. (Eds.) *INTERCHI'93 Conference on Human Factors in Computing Systems. Bridges Between Worlds*. Amsterdam, The Netherlands, 24-29 April 1993, pp. 314-320.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 1996, Vol. 52, (1), pp. 3-50.
- Ingwersen, P. (1992). *Information Retrieval Interaction*. London, UK: Taylor Graham.
- Kuhlthau, C. (1993a). *Seeking meaning. A process approach to library and information services*. New York: Ablex publ.
- Kuhlthau, C. (1993b). A principle of uncertainty for information seeking. *Journal of Documentation*, 1993, Vol. 49, (4), pp. 339-355.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge Series on Human Computer Interaction 9. Cambridge: Cambridge University Press.
- Norman, D. (1988). *Psychology of everyday things*. BasicBooks.
- O'Day, V. and R. Jeffries (1993). Orienting in an Information Landscape. How Information Seekers Get From Here to There. In: Ahslund, S., et al. (Eds.) *INTERCHI'93 Conference on Human Factors in Computing Systems. Bridges Between Worlds*. Amsterdam, The Netherlands, 24-29 April 1993, pp. 438-445.

Preece, J., Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey (1994). *Human-Computer Interaction*. Wokingham, England: Addison Wesley.

Rasmussen, J., A. M. Pejtersen and L. P. Goodstein (1994). *Cognitive systems engineering*. New York: Wiley.

Rouse, W. B. and S. H. Rouse (1984). Human information seeking and design of information systems. *Information Processing & Management*. Vol. 20, No. 1-2, 129-138, 1984.

Saracevic, T. (1996). Relevance reconsidered '96. In: Ingwersen, P. and Pors, N. O. (Eds.) *CoLIS 2. Second International Conference on Conceptions of Library and information Science: Integration in Perspective*, Copenhagen, Denmark. Oct. 13-16, 1996. pp. 201-218. Copenhagen, Denmark: The Royal School of Librarianship. 1996.

Smith, W., B. Hill, J. Long and A. Whitefield. (1997). A design-oriented framework for modelling the planning and control of multiple task work in secretarial office administration. *Behaviour & information technology*, 1997, vol 16, no. 3, pp. 161-183.

Sutcliffe, A. (1997). Task-related information analysis. *Int. J. Human-Computer Studies* 1997 (47) 223-257.

Wilson, T. (1997). Information behaviour: An interdisciplinary perspective. *Information Processing & Management*, Vol. 33, No. 4, pp. 551-572, 1997.