NILS PHARO

The SST Method Schema

A Tool for Analysing Work Task-Based Web Information Search Processes

■

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty of Information Sciences of the University of Tampere,
for public discussion in the Paavo Koli auditorium,
Kanslerinrinne 1,Tampere,
on June 1st, 2002, at 12 o'clock.

Acta Universitatis Tamperensis 871
University of Tampere
Tampere 2002

ACADEMIC DISSERTATION
University of Tampere,
Department of Information Studies
Finland

Distribution

University of Tampere
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 215 6055
Fax +358 3 215 7685
taju@uta.fi
http://granum.uta.fi

Cover design by
Juha Siro

Printed dissertation
Acta Universitatis Tamperensis 871
ISBN 951-44-5354-9
ISSN 1455-1616

Electronic dissertation
Acta Electronica Universitatis Tamperensis 178
ISBN 951-44-5355-7
ISSN 1456-954X
http://acta.uta.fi

Tampereen yliopistopaino Oy Juvenes Print
Tampere 2002

Abstract

In this dissertation, a method schema developed to analyse work task-based Web information search (WIS) processes is presented. The method schema provides a conceptual framework which suggests a set of categories and attributes that play important roles during Web searching and the relationships between these categories. In addition a set of procedures for collecting and analysing data is suggested. The method schema provides a flexible format and is designed in order to provide researchers a strategy for how to perform research on WIS processes. The format makes it possible for researchers to add new features to the conceptual framework in order to describe their problem area as well as new procedures to suit their specific research problems.

Work task-based Web searching can be compared to the processes decision-makers go through when dealing with complex problems. During the process the searcher considers a restricted number of alternatives and consequences, she adjusts her objectives during interaction, she treats her data in a reconstructive way, and WIS processes generated by long-lasting work tasks may be subject to serial analysis and evaluation. The method schema is developed with this view of the searcher's role in the process. In addition to the WIS process itself four major categories have been identified as being part of the interplay with the WIS process; these are the work task, the searcher, the social/organisational environment, and the search task. The work tasks generate search tasks which in turn initiates WIS processes. The WIS process in turn consists of search situations and search transitions reflecting the searchers' interaction with information and surrogate resources respectively. During a WIS process there is a dynamic interplay between all five categories. The method schema is designed to develop methods to capture this interplay.

The method schema was developed using a combination of inductive and deductive thinking. Studies of literature on information behaviour provided a basis for our understanding of factors that influence information searching. A set of data of real WIS processes, including video recordings of processes accompanied by recordings of the searchers talking aloud, was collected. A subset of the data was used for a pilot study in order to identify important characteristics of the processes. Grounded theory was used to develop the conceptual framework.

The method schema was justified by the development of a method for analysing the WIS processes of a group of students. Data on 9 WIS processes were collected, comprising 652 situations and transitions and more than 12 hours of video recordings. A set of empirical research questions was investigated. The results of the analysis showed that work task goals directly influenced the relevance level used by the students. It also indicated that the work task stage directly influences the length of situations and transitions. Most importantly the study shows that it is possible to use the method schema for analysing relevant problems of WIS processes.

Acknowledgements

It is with a certain amount of joy that I present my doctoral dissertation. It is the results of approximately 4 years of work which would not have been completed without the help of a group of people and organisations:

The Norwegian Research Council's Programme for Library Research has financed the project, and I am thankful for the practical administration provided by the Programme's secretary Ruth Jensen. Professor Ragnar Audunson could have been thanked in several of the paragraphs, but I choose to use the first opportunity. Ragnar has been the leader of the Research Programme and been very important in initialising research on library and information science in Norway and he has been very encouraging in matters concerning my research. The other Ragnar, Ragnar Nordlie has been a very valuable discussion companion in all stages of my project. Ragnar was the supervisor for my master thesis and we have had neighbouring offices at the Oslo University College throughout the project work. I would like to thank him for the stimulating professional as well as the social time we have shared.

The Faculty of Journalism, Library and Information Science at the Oslo University College has provided me with an office and an environment that has made it possible to perform my studies. My fellow doctoral students at the Faculty have provided a good environment for discussions.

The research environment for studies on information behaviour is small in Norway, it was therefore a great pleasure for me to be enrolled at the Department of Information Studies at the University of Tampere. The NordIS-NET network has been another important institution. The network has arranged and financed research courses for doctoral students from the whole Nordic region. This has made it possible for me to meet and discuss with distinguished scholars and fellow students from the Nordic countries as well as the rest of the world. I believe NordiIS-NET has been very successful in creating a more united research environment in the region and hopefully the network will continue on this path.

Professor Kalervo Järvelin has been my supervisor. The fact that we have been located in two different geographic settings during most of the research project necessitated intense supervision periods whenever we could meet supplemented with the use of e-mail. This, in my view, has worked very well. Kal is very creative and has offered me very high-quality feedback. He has provided some very good ideas about how to improve my work, for which I am deeply thankful.

Professor Pertti Vakkari has been another good discussion partner at the University of Tampere.

I am very grateful for the warmth I have met at the Department of Information Studies and for the professional interest you have had in my work. I would especially like to thank the doctoral students who I have had the pleasure to meet and who have made me feel very welcome in Tampere, a special thanks to Kai Halttunen who welcomed me in his

home! The FIRE group has been an encouraging forum for discussion and it has raised valuable questions for discussion. I would also like to thank the Department's adminstration, especially Mirja Björk, Leena Lahti, and Nina Rainevuo who has helped in all practical matters.

Professors Peter Ingwersen (Royal School of Library and Information Science) and Raya Fidel (University of Washington) has been the assessors of my work and has provided good comments about how to improve it.

I also thank Eric Deverill for helping me in improving the language and Kjetil Iseli and Kamran Bagheri for technical support.

My family and friends also deserves a "thank you" – thank you for your support and encouragement!

Last, but not at all least, I want to thank my wife Kristine and our son Tor. Kristine has been a tremendous support for me, for longer periods she had to be a "sole parent" and I know it has not been all that easy. Now I believe this situation will improve.

Oslo, 2 May 2002

Nils Pharo

1 Introduction 10
1.1 User centred tradition of information behaviour 11
1.2 Central concepts 12
1.3 Research problem 15
1.4 The structure of the dissertation 18
2 Web and web interaction 19
2.1 Web resource types 20
2.2 The Web compared to other information systems 22
2.3 Previous studies of Web interaction 24
2.3.1 Demographic studies 24
2.3.2 Usability studies 24
2.3.3 Library and information science based studies 25
2.4 Web interaction research needs 28
2.5 The strategy of disjointed incrementalism 29
2.6 The demands for developing the method schema 32
3 Evaluation of previous models of information seeking and user-centred IR ..34
3.1 Evaluation of models of information behaviour 34
3.1.1 Evaluation criteria 35
3.1.2 Models of information seeking behaviour 38
3.1.3 Models of information searching behaviour 49
3.2 A discussion of the existing information behaviour models 57
4 Designing the method – problem statement and procedure 64
4.1 Ill-structured problems need flexible methods 64
4.2 The development of the method 66
4.3 The sketch of the schema 69
4.3.1 The sketch of a framework 69
4.3.2 Elements of the procedure 71
4.4 The pilot study 73
4.4.1 Data collection 73
4.4.2 Introducing grounded theory 75
4.4.3 Grounded theory applied on web search sessions 76
4.4.4 Real time data analysis 77
4.4.5 Real time analysis results 79
4.4.6 Analysis of transcriptions of video recordings 82
4.4.7 Results of transcriptions analysis 85
4.4.8 Story analysis 88
4.4.9 Summary of the pilot study's methodological parts 94
4.5 The domain description 95
4.5.1 Work task 95

4.5.2 The searcher 97
4.5.3 Social and organisational environment 100
4.5.4 Search task 100
4.5.5 The search process 102
4.5.6 Relationships 107
4.5.7 Summary of the method schema's domain 111
4.6 Procedure of the method schema 112
4.6.1 Data collection procedures 112
4.6.2 Data analysis 115
4.6.3 Summary of data collection and analysis 117
4.7 Evaluation of the method schema's domain (and procedure) 117
4.8 Summary 123
5 An empirical study of Web users – justification of our method schema 124
5.1 Demands to a justification 124
5.2 Empirical research questions 124
5.3 Collecting the data 126
5.3.1 Student characteristics 127
5.3.2 The questionnaire 127
5.3.3 Pre-search interviews 132
5.3.4 Observation and video recording 133
5.3.5 Post-search interviews 134
5.4 Creating stories 134
5.4.1 Validity of story creation 135
5.4.2 How to classify categories and attributes 136
5.4.3 How to classify relationships? 137
5.4.4 Coding 138
5.4.5 Interrelationships 139
5.4.6 The action – resource type relationship 139
5.4.7 Some examples 139
5.5 Analysing stories 141
5.5.1 Situations and transitions 142
5.5.2 Empirical research questions 147
5.5.3 Influence of work task goals on WIS processes 147
5.5.4 Influence of search task strategies on WIS processes 151
5.5.5 Influence of work task knowledge and search knowledge on WIS processes 158
5.5.6 Influence of work task stages on WIS processes? 164
5.6 Summary of the empirical study 165
6 Summary and conclusions 169
6.1 Goal and terminology 169
6.2 Summary 170

6.3 The need for methods for information behaviour research 172
6.4 Justifying the SST method schema 173
6.5 Describing and prescribing information behaviour 175
6.6 The flexibility of a method schema 176
6.7 The SST model and qualitative/quantitative research 178
6.8 Problem types for which to use the SST method schema 179
6.8.1 System-focus 180
6.8.2 Searcher-focus 180
6.9 The SST model applied on other information systems 182
6.10 Information seeking and searching – a need to combine efforts 183
6.11 Limitations of this study 185
6.12 Conclusions 186
References 187
Appendix 1: Questionnaire 196
Appendix 2: Interview guide 202
Appendix 3: Classification rules for categories and attributes 203
Appendix 4: Transcript of session no. 7. 212
Appendix 5: Story representing search session 7 226

1 Introduction

The Web has been a focus of interest among many scientific disciplines; the library and information science (LIS) (and computer science) community has focused on information seeking and retrieval issues related to the Web. The efficiency of search engines (e.g. Gordon & Pathak, 1999) and in particular query term statistics (e.g. Silverstein et al., 1999; Jansen, Spink & Saracevic, 2000) have been the topic of several authors. These studies, however, tell us little about the nature of the searchers using them and why they have decided to use that particular resource. Neither do we know much about the purpose of their searching. For example why do they enter those particular search terms? Previous research on Web searching has not taken into consideration the characteristics of the Web-searcher interaction and how this affects the search process. In the present study we wish to explore such a perspective in order to create a method that makes it possible to understand the interplay of the key elements in Web searching during Web information search (WIS) processes.

The claims from the literature on information seeking (e.g. Wilson, 1981; Wilson & Walsh, 1996; Ellis, 1989; Ellis, Cox & Hall, 1993; Kuhlthau, 1991; 1993) are that searchers are characterised by their motivation and previous search experiences as well as the work task they intend to perform. The search task is the manifestation of how the searcher intends to solve (parts of) his work task with the help of an information system. We believe that it is reasonable and necessary to take such characteristics of information seeking into consideration when studying WIS processes.

Although it is generally agreed that work tasks affect information behaviour, we find few studies which investigate this on a deeper level (Vakkari, 2003). There is very little theoretical analysis (Byström & Järvelin, 1995; Vakkari, 2001) and even less empirical work done to support the theory. It may be easy to agree that the work task affects information needs and thus the search topics. If we are satisfied with agreement on this level, we have, however, made much argument out of almost nothing and can abandon the work task as a topic of information behaviour research in favour of focusing on the *search* task. Before dismissing work task entirely as an object of study, however, there are a few questions that need to be investigated:

- What dimensions of work tasks affect (Web) information searching? What work task dimensions exist?
- What parts of the WIS process do the work tasks affect? Which dimensions affect which parts of the WIS process?

- How do these dimensions affect, and interact with, other factors? How are work task-related analyses used, explicitly or implicitly, during WIS processes?

If we consider work task an important factor for Web information searching, then we should treat it analytically and analyse its effects. This has been a basic premise for the work on the present project.

We want to focus on what happens during actual Web-searcher interaction. Here there are lessons to be learned from the literature focusing on user-centred information retrieval (IR) (e.g. Ingwersen, 1992; 1996; Belkin, Marchetti & Cool, 1993; Belkin et al., 1995; Saracevic, 1996a; and Spink, 1997). It seems relevant to apply concepts relating to, for example, interaction techniques, search reformulation, and relevance judgement to our framework and we are in particular interested in addressing how reflections during the search process affect the searcher.

1.1 User centred tradition of information behaviour

Our work is related to the so-called "cognitive viewpoint" in information retrieval (e.g. Belkin 1990, Ingwersen 1992; 1996; 1999). This tradition within LIS is founded on De Mey's definition that any information processing "whether perceptual or symbolic is mediated by a system of categories or concepts which, for the information processing device, are a model of its world" (e.g. in De Mey, 1982, p. 48). This viewpoint runs hand in hand with the user-oriented studies of information seeking behaviour. Traditionally information retrieval (IR) research has focused on the retrieval systems and the objectives have been to develop algorithms that are better at representing and matching document and query terms.

The cognitive viewpoint has been criticised for being too individual-oriented in the sense that it focuses on interaction within entities in the human mind (Frohmann, 1992), a criticism that has been met by Ingwersen, who claims that this is a misinterpretation of it "being identical to 'mentalism' or 'cognitivism'" (Ingwersen, 1993, p. 61). Ingwersen, in particular, has been advocating a direction of the cognitive viewpoint that also takes into account the problem solvers' social and organisational environment and how it influences information searching. The term "cognitive" is thus not the best to denote this tradition within IR.

For us it is of less importance what to call the tradition. We try to focus on the user working within an environment where she[1] is forced to search for information in order to solve her problems or work tasks. There are many aspects that come into play during information behaviour and it is clearly insufficient to solely focus on the information system-side aspects.

1.2 Central concepts

We intend to create a method schema, which in turn involves representing the class of problems applicable for the method in the form of a conceptual framework. In order to clarify the purpose of our study we have found it necessary to specify what we mean by some central concepts.

A *method* is, according to Bunge (1967, p. 8) a procedure for handling a set of problems. Methods can be categorised as "quantitative", which is, for example, the case for various statistical ways of data handling, or "qualitative", which may be exemplified by grounded theory (Glaser & Strauss, 1967). We have used the term to focus on methods in a broader way. It can be used to treat problems on different levels of generality.

According to Newell (1969) a method consists of the following three parts: (1) a problem statement or domain, (2) a procedure, and (3) a justification. In more detail we may describe the three elements of a method in the following way:

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

The *procedure* is what we in daily terms would call the method itself. It is "what delivers the solution to the problem" (Newell, 1969, p. 370). The procedure refers to the operations that should be applied to trigger its execution. It may, for example, include specification of the kinds of data needed to identify a certain effect.

Supplying arguments in order to make it rational to believe that it works constitutes the *justification* of the method. This applies to testing the method with the specified data to show that it delivers the solution to the problem. When dealing with problems that demand well-structured methods with a low generality and high power (Newell, 1969) the

[1] We have deliberately and systematically used the feminine gender (*she*) when referring to the searcher; this has been done because the majority of searchers observed in the study were women.

justification of the method will consist of supplying the required data and showing the procedure's success[2]. The method can then be compared to competing ones by comparing different properties such as the following, suggested by Järvelin and Wilson (2002); the simplicity, accuracy, scope, systematic power, explanatory power, fruitfulness, and validity.

A *method schema* takes the format of a method, but it contains unspecified components that need to be specified if it is to reach the level of a method. According to Eloranta (1979, p. 125) it is any representation defined for one or more methods, where one or more aspects of the method have been left uninterpreted and represented only through their plain name, and where some aspects of the methods may have been left out (even lacking their naming). In other words a method schema is an abstract representation of one or more methods – a generic model. The difference between a method and a method schema can be said to be a continuum of generality.

A *conceptual framework* is, according to Engelbart (1962), the specification of:

- The important factors of the problem[3] to be studied
- The relationships of the factors that are recognised
- What kinds of changes in the factors or their relationships affect the functioning of the problem, and how these changes affect the problem
- Promising or fruitful research goals and methods

A conceptual framework can thus be said to integrate a set of theories in order to handle research problems within a specific domain. We use the conceptual framework to map the domain of our method. The procedure of the method is discussed separately from the domain, although Engelbart conceptually treats it as a part of the framework.

Methodology is the term used for the study of methods. The guidelines presented above stating how to create a method (state the problem, explain how it may be solved, and justify it by testing) thus exemplifies methodology when applied in the definition of a method. Methodology, on the other hand, can also be defined as "a body of methods, procedures, working concepts, rules, and postulates employed by a science, art or discipline" (Gove, 1971). This second meaning corresponds to a certain extent with our

[2] Newell (1969, p. 371) uses the quadratic equation as his example. It is easy to see that the justification of a problem involving searchers that reflect on their own role in the searching process will not be that "simple".
[3] In the original text Engelbart (1962) uses the term "system" (referring to human intellectual effectiveness) instead of "problem". We prefer to use the latter.

understanding of the conceptual framework. Nevertheless we have chosen to use the latter term to denote the domain of our method.

A scientific *theory* states a relationship between two or more phenomena. A theory is narrower than a conceptual framework and it is less fundamental. A theory should be falsifiable while a framework is more related to a paradigm or a research program as Kuhn (1996) and Lakatos (1970), respectively, have named the set of properties of a scientific area that provides the background against which all research in the area is carried out[4]. Tests of a theory are made by testing hypotheses that explicitly state relationships between phenomena under study.

The term *model* is used to define a representation of a problem. We choose to view it as a more general concept than scientific theories and conceptual frameworks in order to use it as a collective term for describing different approaches to creating understanding of information behaviour.

Information is used in a rather limited way to denote that which reduces a person's uncertainty (Wersig, 1971). Something cannot be said to be information if it is not the result of interpretation of data by the receiver.

Paradoxically the term "information" may be used in the text as part of compound terms, such as "information seeking", "information searching", and "information retrieval", which are defined below. In such settings the term loses the stringency of our definition, since it will more or less take a meaning synonymous to "data". Ingwersen (1992, p. 32) proposes the use of "potential information" to denote data that has yet to be perceived by a recipient. We choose to follow Ingwersen's advice.

Information seeking is used to describe studies of how people react when they experience a need for information in order to perform a work task. It involves all kinds of cognitive, emotional, and physical reactions which this evokes and includes their attempts at clarifying the need, expressing it, selecting among possible sources to satisfy it, interaction with potential sources, and refining the result of the interaction processes.

[4] Note that a paradigm (Kuhn 1996) also represents beliefs and values for the research community and thus can be viewed as a social construction. Developing our framework we have not focused on such a dimension. Tuominen (2001) discusses *meta*-theories and states that these often are implicit, unarticulated, and socially shared: there always is community reproducing the discourse and sharing the understanding. We believe the same aspects are characteristics of paradigms.

Information searching constitutes the user-source interaction part of information seeking. Information searching may involve persons interacting with any kind of information system and sources, including other persons.

Information retrieval (IR) is used to denote interaction between persons and (computerised) information systems. Traditionally IR-research has focused on the system-side of interaction, although a user-centred tradition of IR has evolved in the last 20 years. We prefer to use the term information searching, which in our view is more inclusive, in particular because it does not have such a strong relationship with computerised information systems and sources.

The *World Wide Web* is a distributed, multimedia, hypertext system; "distributed" because the content is saved on computers all over the world; "multimedia" means that it consists of texts, graphics, sound, video etc; and "hypertext" means that the content is structured non-linearly, i.e. it can be accessed in a largely user-defined sequence. We will often use the short form "Web" with capital W when speaking about the system as a whole, "web" is used when it forms part of a compound term, such as "web interaction", "web resource" etc.

A *web page* is a web client program's (e.g. Netscape Navigator, Opera, or Internet Explorer) viewing format of an HTML[5]-file.

A *web resource* is a collection of pages designed and structured to constitute a whole. It will often be published by a single organisation and have a common co-ordinator.

By *web interaction* we mean a person's use of the World Wide Web for searching for information.

1.3 Research problem

The main goal of the study is to develop a method schema for analysing Web information search (WIS) processes. It should take into consideration central concepts identified in both information seeking and user-centred IR literature. We have used the methodological directions given by Newell (1969) to represent the domain, procedure, and justification of the schema. Due to the apparent complexity of the problem we have found that it would be suitable to develop a conceptual framework according to the specifications of Engelbart (1962). We have developed a method *schema* and not a method

[5] Or related file types such as XML

because we believe that it is more realistic to keep it at this level and open for future researchers to design the appropriate method from this schema in accordance with their own research problems. To design a general method would necessitate resources we have not been equipped with and we believe that the format we have chosen can serve as a valuable tool. We also believe it is sound to be methodologically clear and do not wish to blur the discipline with a tool that gives an expression of being something it is not. As the reader shall later see we have used the schema to specify a method for a particular setting.

We present the research problem as follows:

1. The *domain of the method schema*. To describe the problem it will be necessary to map the central concepts and their relationships in a format that makes it less complex. We will use the literature within information seeking and IR to create a *conceptual framework* that presents the problem in a comprehensible way. The framework should:
   - be suitable for analysing search processes in the Web which may be characterised by a multitude of sources, information and document types, searching and browsing tools all available through a single browser;
   - represent characteristics of the WIS process as well as external factors which have an impact on the WIS process, for example factors related to the searcher, her work task, search tasks, etc.;
   - provide an effective level of abstraction - pruning out unnecessary details while retaining an analysable view on the web interaction. Therefore it should suggest units of analysis for the process, which should be of a reasonable size and where the interplay of the units can be identified;
   - suggest attributes or dimensions of the units that capture the relevant features for understanding and explanation of the interplay;
   - suggest classification of the units or their attributes that allow for asking fruitful questions or detecting any connections between them.

2. The *procedure of the method schema*. We present the operations that need to be performed in order to analyse WIS processes and the factors that affect it. This involves, in addition to general methods of research (such as design of projects, rules of writing etc.), identification of in particular:
   - methods for data collection. What methods should be used in order to collect data that represent the different characteristics of WIS processes and affecting factors;

- methods for data analysis. How we can use the collected data to identify the interplay in the WIS process that occurs both between the units as well as within the units.

3. The *justification of the method schema*. The schema can be justified by showing its ability to solve designated research problems of the domain, i.e. by using it to generate a method for describing actual WIS processes. In addition it should be tested against a set of formal requirements put forward by Bunge (1967). Bunge developed the criteria for evaluating scientific theories, but it should also be applicable to methods and method schemas using our broad definition of the concept. This is because the method schema's domain is formulated as a conceptual framework, which, in turn, is seen as a broader representation of a theory/a set of theories (cf. the definitions above).

   The method should support systematisation of knowledge by:

   - compatibility with previously well-tested knowledge;
   - integration of formerly separate parts of knowledge;
   - generalisation and explanation of lower abstraction level knowledge (observations, data) through higher level constructs;
   - explanation of facts through systems of hypotheses which entail the facts;
   - expanding knowledge by deducing new propositions based on selected starting points and collected information;
   - improving the testability of hypotheses through the control context provided by systems of hypotheses.

   The method should guide research by:

   - pointing out fruitful problems;
   - proposing the collection of data which nobody would understand how to collect without the theory;
   - proposing new lines of research.

   The method should be able to map (parts of) reality by:

   - representing or modelling the objects (and relationships) of the relevant parts of reality instead of plain summarisation of data;
   - suggesting tools for producing new data.

1.4 The structure of the dissertation

In Chapter 2 we examine the nature of Web interaction and point out what has been done in previous research on Web information behaviour within LIS. We identify important characteristics of the Web and discuss how it differs from other information systems. This chapter provides the background for the present study.

Chapter 3 evaluates previous models of information behaviour in order to identify major factors that seem to play a role when users try to solve problems or execute work tasks. We provide a set of criteria to evaluate existing models in order to clarify their methodological value. Another set of criteria is used to evaluate the existing models' applicability to analysing Web information search processes.

Chapter 4 presents our method schema, which is called the search situation transition (or SST) method schema. It discusses the methods we have used to develop it (i.e. the *meta*-methods) and the different steps the work has gone through (Subsection 4.1-4.5). We then describe the schema's domain description (Subsection 4.6) and its procedure (4.7). In the chapter's final subsection (4.8) we start the justification by evaluating the schema according to the formal criteria introduced in Chapter 3.

Chapter 5 contains the justification of our schema. We present an empirical case study and show how the schema can be used to generate a method for examining a set of empirical research questions (Subsection 5.2). Subsections 5.3 and 5.4 contain descriptions of the techniques we have used to collect and analyse data (the procedure). In Subsection 5.5 and 5.6 we present the results and conclusions of the empirical study.

Chapter 6 is the final chapter and contains discussion about possible uses of the method schema. It contains an extensive summary of the dissertation as well as our main conclusions.

2 Web and web interaction

The World Wide Web was created at the European Laboratory for Particle Physics – CERN – in the early 1990s (the Web is itself the best source on its history, e.g. in Cailliau, 1995). The original intentions behind the Web were to encourage collaborative work across physical borders. The hypertext format makes it possible to incorporate texts within each other using links. Thus a researcher in the USA may write comments, examples and/or explanations to a text put out by a Norwegian colleague on a Web-server that they both have access to. This made it possible to leave the original text untouched while at the same time opening up for incorporation of co-workers' comments etc.

What originated as an academic tool has in only a few years developed to become a virtual world consisting of digital equivalents to many physical world phenomena. Lawrence & Giles (1998) estimated the minimum size of the indexable Web to be 320 million pages in December 1997. According to a survey performed by Brightplanet the deep Web, i.e. the "ordinary" flat Web plus databases that are not indexable by search engines, consists of approximately 550 billion pages (Bergman, 2000).

Lower computer prices and the introduction of graphical interface browsers created a new potential Internet public. Simultaneously the non-academic actors began to see the potential in Web publishing. Many were commercial organisations that had some service or product to sell, but also local and governmental institutions discovered that the Web could serve their interest as an information channel. In the academic sector there were still many institutions that had not got connected to the Web. The number of academic institutions on the Web has continued to rise in the years after the introduction of graphical browsers. One particularly important group of Web actors is the Internet Service Providers (ISPs). The ISPs make their living out of selling access to the Internet and often make agreements with computer distributors so that those who purchase a computer usually are offered an Internet account from one of the ISPs.

The Web actors now constitute a more heterogeneous group than ever which is reflected in the variety of web resources available. The technological development has made it possible to include videos and sounds in several formats with the help of additional software accessible via the browser, the so-called plug-ins. The use of databases has made it possible to generate Web pages by combining elements from separate files, thus creating more dynamic resources.

In our study we have focused on literature concerning the Web from an LIS point of view in general and more specifically studies that focus on Web information behaviour. To a certain extent we have supplemented this with studies of Web usability.

2.1 Web resource types

In the beginning most resources were "copies" of physical world documents like scientific articles, student handbooks and fictional literature without copyright restrictions, which were more or less directly transferred to hypertext. Now distinct Web genres have been developed and old document types have incorporated features developed within the new genres. The most important feature is interactivity, i.e. the possibility to interact with congenial people across the world in real time. So-called chat rooms have been identified to be very attractive to Web users. Sites offering such services may, according to Hof, Browder and Helstrom (1997), expect up to 50 % rise in traffic. These sites can be prepared to keep their users for a lot longer time than sites not offering chatting, the latter of course having a large impact for those trying to earn money by selling advertising space.

Via the Web it is possible to access information directly or indirectly, e.g. bibliographical databases are made available through the net, which makes it possible to locate and order books or journal articles on-line and have it sent by regular mail. Some journals are also available in full text on the Web making it possible to read it on screen or print it out. Some of the journals are freely available while for others some kind of fee structure may be present.

Murray Turoff (1997) has pointed out that the computer has made it possible for us to make systems not necessarily mirroring the physical world:

> "With computers, we have a commercially available and popular product that allows virtual environments to be accessed and used from anywhere and at any level, from individual applications to world-wide networks. We "implement" systems that can be based on any possible metaphor, without foundation in validated theory or scientific evidence." (p. 39)

An, in our opinion, even more interesting point he makes is to state that "these virtual computer systems can become formal models and representations, or validation vehicles that can be used as templates *not only to describe the real world but to prescribe it*" (p. 39) (this author's emphasis).

If we consider the Web in the light of Turoff's article we may look at it as consisting of a myriad of virtual environments, where only fantasy sets a limit to the resources possible being made available on and through the Web. The Dublin Core project (Dublin Core metadata initiative, 2002) is an attempt at creating cataloguing rules for the Web. The project has used the term "document-like-objects", or DLOs to describe Web resources[6].

The DC project has come up with 15 different elements to describe Web resources. Of these the "type" element is used to describe

> "the category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, Type should be selected from an enumerated list that is currently under development in the workshop series" (Dublin Core Metadata Element Set, 1998)

The work being performed by the DC project will be interesting to follow. They have evolved from being primarily paper world-oriented towards integrating resource types like "ambient sounds", "executable software" and "interactive games". The present recommendation (DCMI Type Vocabulary, 2000) suggests 9 general resource types: collection, dataset, event, image, interactive resource, service, software, sound, and text. Apart from establishing a format for cataloguing Web resources the DC project may have other consequences. The project's value as a forum to illuminate Web properties should not be underestimated. Of particular interest is the mailing list maintained by the project, which is a meeting place for librarians and computer scientists involved in knowledge organisation on the Web.

Ben Shneiderman (1997) discusses Web genres and how these influence Web designers. He suggests four ways of categorising Web sites; by the originator's identity; by the number of web pages or amount of information that is accessible; by the originator's goal; and by the measure of success. These are important facets of Web resources, but we believe a finer granularity is needed to treat the resources from a knowledge organisational viewpoint.

The Web terminology is in our opinion quite blurred; the following specifications are our own. We have chosen to use Web "page" to denote the constellation of text, pictures, videos and sounds that emerges as the result of activating a URL in a browser. A Web

[6] In our opinion we should be very careful with treating everything on the Web as documents. It would be a mistake to e.g. view a virtual museum as a document, on the one hand because its physical mirror is not regarded as a document, on the other because it may possess properties that are realised specifically through digitalisation, e.g. IR possibilities or interactive chat services.

"resource" is usually a collection of pages designed after a standard set of principles, but it may also consist of a single page. There will usually be a co-ordinator for each resource, but the different pages may be scattered on servers world-wide. The term "site" is used about an organisation's collection of Web pages. A site may include several resources and more sites may reside on a single server.

As an information environment the Web constitutes an interesting research object. The fact that we may now use a web client program to access information services that we previously got in touch with individually on the shelves, by telephone, in meetings, or by proprietary software, is a way towards fulfilment of the cliché "information at your fingertips". This, of course, seriously challenges the intermediary role of the librarian, who may need to use her knowledge of users to design user-friendly Web information systems rather than to perform the actual searching for the users.

The chaotic information environment growing out of this contains great potential for helping people in problem solving, but effective exploitation of the Web faces several challenges of a technological and structural nature.

2.2 The Web compared to other information systems

Researchers on information systems have usually focused on very structured systems, i.e. systems where the potential information is organised in a predefined set of categories. Examples of such systems are library card catalogues, bibliographic databases, and citation indices. The Web is, on the other hand, a very unstructured information system. Most users do consider the Web a single system; they seldom look at more than four to five pages in any single resource before continuing somewhere else (Nielsen, 2000, p. 189). Such a user-centred view is necessary to take when designing usable web pages and resources since this requires the ability to create pages that keep the user's attention.

There are two fundamentally different approaches one might take when dealing with the Web as a resource for users' problem solving:

- to view the Web as a collection of individual information systems where the interaction with each systems should be treated separately;
- to look at the Web as one information system where the transitions between different sources constitute a significant part of the interaction process.

The first approach is by far the most usual. It makes it possible to compare the performance of web-based and non-web-based systems - making it a system-oriented approach. The large studies of query term statistics performed by Silverstein and his colleagues (1999) and Jansen, Spink, and Saracevic (2000) are good examples of the first approach. If the second option is chosen the user will be in focus, but it necessarily makes it much more complex to identify and compare source characteristics. The study by Catledge and Pitkow (1995) exemplifies the second approach; the researchers collected Web client program logs of searchers which included searchers changing between different resources during their sessions.

This project is based on a holistic and process-oriented view of web interaction. This implies that we see the user as interacting with the whole Web and we look at the interaction as a continuum where the user builds her knowledge about the Web, i.e. we take into account that the user's Web model is dynamic.

It is, however, necessary to understand site-dependent features and how users' experience with different web sites influence their interaction. Users may, for example, learn to successfully generate relevant information from one site with a particular technique and therefore use similar techniques in other sites expecting this to be similarly efficient.

In Web interaction users may typically drift between different *types* of resources as well as different resources. When users interact with bibliographic databases they will behave differently from when they read a full-text article in a newspaper or if they try to find the email address of an old friend. The Web search *process* is a process of change between different resources and with different means. This is a property of search behaviour that is very Web specific, which thus constitutes an important object to study. From previous studies of information behaviour we know that it depends on many different factors other than the resource types themselves, such as factors relating to e.g. the searcher's work task (Byström, 1999), personal characteristics like education (e.g. Ellis, 1989; Ellis, Cox & Hall, 1993), search knowledge (e.g. Ingwersen, 1992), and subject knowledge (e.g. Marchionini, Lin & Dwiggins, 1990), organisational factors (e.g. Ginman, 1987), as well as factors that come into effect in the search process itself – for example the actions performed (Belkin, Marchetti & Cool, 1993), relevance judgements (Saracevic, 1996b) and the accumulated results (Bates, 1989). Some of these factors may influence interaction with any kind of information systems, including the Web, but there may exist others that are Web specific. It is necessary to take such properties into consideration when one wants to understand or make sense of the Web search process.

2.3 Previous studies of Web interaction

A number of studies of Web use have been performed (Jansen & Pooch, 2001). These include large-scale surveys focusing on users' backgrounds and intentions when using the Web, i.e. their Web information seeking behaviour; examination of the use of individual Web resources and/or sites; and studies of navigation across multiple Web resources. The data are collected in various ways depending on what aspect of web behaviour they seek to reveal; questionnaires, interviews, log analysis, and observation are common data collection methods.

2.3.1 Demographic studies

Major demographic studies have been performed by for example CommerceNet (Internet Demographics and eCommerce Statistics, 2000) in co-operation with Nielsen Media Research (Interactive Services, 2000). In addition, the web-based surveys run by the GVU centre at Georgia Institute of Technology (GVU Center's WWW User Surveys, 2000) attract web users to input demographic data in an online environment. They are however of little interest when it comes to identifying actual Web interaction although they might tell us something about the kinds of services that different groups of users are interested in.

2.3.2 Usability studies

On the opposite level of the web survey scale we find the usability studies. According to Jakob Nielsen (1995) usability is associated with the following five attributes: easy to learn; efficient to use; easy to remember; few errors; and pleasant to use. In Nielsen (1998) he divides Web usability problems into two categories:

- Site-level usability: home page; information architecture, navigation, and search; linking strategy; internally vs. externally focused design; overall writing style; page templates, layout, and site-wide design standards; graphical language and commonly used icons;
- Page-level usability: specific issues related to the individual pages: understandability of headlines, links, and explanations; intuitiveness of forms and error messages; inclusion or exclusion of specific information; individual graphics and icons.

There are many web resources devoted to HCI and usability. Of these Nielsen's "useit.com[7]" is a good guide in writing for the Web. There is little doubt that users may face problems when interacting with badly designed Web pages and resources. Whether there are errors that systematically hinder them from fulfilling their tasks depends heavily on the tasks and searchers.

To evaluate Web browser design Chen et al. (1997) performed a small-scale experimental study of users performing specific search and non-specific browsing. For browsing tasks the mean path length was significantly higher than for specific search tasks. The survey showed no significant difference in use of basic navigation and browsing functions, for instance, following links and using the browser's backtracking mechanism which were the preferred functions within this group. For specific search purposes there was however a main effect in the use of functions initiating new paths, for instance, using links to search engines or submitting keywords to a search engine. These functions were – not surprisingly – used more often for specific search tasks.

Cockburn and Jones (1996) studied usability problems of Web client programs, or browsers, in particular focusing on their navigation facilities. They discovered discrepancies between the users' mental model of the history list, which also influences the use of browsers' forward and back buttons, and the actual list being available.

Richmond (1996) has looked at the HCI and usability features that entice users to shop in web environments and points out success factors such as animated graphics and the ability to chat with others. To make payment accounts available beforehand was deemed to be the most important factor to increase the amount of money spent during online shopping.

Although usability studies focus most of their attention on how one may improve the system side of interaction, the findings done by usability researchers could be used to explain why web interaction might follow certain paths.

2.3.3 Library and information science based studies

The majority of LIS-based studies of Web interaction are focused on single sites and based on server log analysis. The major weaknesses of these studies are that they only catch a glimpse of the users' Web interaction and that we know nothing about the user, his/her goals, strategies, and motivations. On the other hand log analysis is an easy way

[7] http://www.useit.com/

of getting hold of data that can be treated with quantitative methods and we can use it to get statistically significant data about users' choice of terms and use of syntax for querying search systems.

Jansen, Spink, and Saracevic (e.g. 2000) have analysed more than 50000 queries in Excite's query log and found that users use few terms when searching the database (2.21 per query). Compared to findings in studies of other information systems this is surprisingly low, since such studies have found the mean number of terms per query ranged from 7 to 15. It should be noted, and so it is by the authors, that the users of individual information systems may be of a very different kind (consisting of e.g. IR specialists) than the general Internet population (p. 216). When removing identical queries it was shown that the users on average entered 1.6 queries per session. This indicates that the users spend little effort per task in a single search engine although the paper says nothing about whether users search for different topics during a session, i.e. we do not know if they tried to solve more than one task in one session. The survey also shows that only approximately 5 % of the users use advanced search features like the Boolean AND-operator (*very few* use OR and AND NOT) and relevance feedback (the latter is used in 5 % of all *queries*). A third important point is the examination of search results. Only 20 % of the users looked beyond the first two results pages. On average each user looked at 2,35 pages. What would have been more interesting is to identify the number of pages accessed via the result lists.

Silverstein and colleagues (1999) have performed a similar analysis of a much larger data set - approximately 1 billion requests, or about 575 million non-empty queries (p. 8) - from Alta Vista. Their findings support the notion that Web users behave differently from users of traditional IR systems - they use few query terms, investigate only a small portion of the result list, and seldom modify queries. It is however impossible to tell what the situation would have been like if the search engines had similar response times and the same features that professional IR systems have. If one wishes to compare differences in use between Web search engines and traditional IR systems, one should take into account both the users, the system and the intermediary, i.e. different human computer interaction (HCI) dependencies like bandwidth, features of the client program, etc. To get such information it is necessary to study interaction from the user side.

There are surprisingly few studies that have focused on the user, or client side, of Web transactions (Catledge & Pitkow, 1995; Wang & Tenopir, 1998; Wang, Hawk & Tenopir, 2000). There are, however, a few studies that have focused on children's and high school students' use of the Web to solve assigned specific search tasks (e.g. Fidel et al, 1999;

Large, Beheshti & Moukdad, 1999; Bilal, 2000; 2001). Lots of user-centred surveys have, on the other hand, been performed with other hypertext systems (e.g. Shneiderman et al., 1989; Marchionini, Lin & Dwiggins, 1990; Rada & Murphy, 1992; Qui, 1993a-b).

An interesting survey was done at the Georgia Institute of Technology (Catledge & Pitkow 1995). Although not performed within an LIS setting we have included it for its originality. 107 persons belonging to the Institute agreed to have their client logs captured over a period of three weeks. The client logs contained the URL of the users' current and target page, as well as information on the technique they used to access the target. The data were analysed to compute path lengths and frequency of paths. A so-called Pattern Detection Module (PDM) algorithm was used and three kinds of Web users were found:

- Serendipitous browsers, i.e. users who avoided repeating long sequences
- General purpose browsers, i.e. users performing as expected. These users have a 25 % chance of repeating complex navigation sequences
- Searchers, i.e. users who repeat short sequences infrequently, but long navigational sequences often

The survey also gives some insight into which techniques and tools are being used to browse the Web. They found that in 93 % of the cases following links (52 %) and using the back button (41 %) was the method being used to access Web pages.

Since the data were collected from the log files of the users it was not possible to measure the success of each session. The categorisation of different Web users is however interesting to note. It shows that different people evoke different Web search strategies.

Wang and her colleagues (Wang & Tenopir, 1998; Wang, Hawk & Tenopir, 2000) are interested in the cognitive styles and affective states of web searchers. One conclusion they make is that many users develop a general mental model that covers all Web search systems. Thus they may for example use the same syntax in different systems, "but there was little evidence that users changed their mental models after a few failed trials with no messages or clues from the system" (Wang, Hawk & Tenopir, 2000, p. 243). They also found that there is an advanced group of Web users who use advanced features of the search systems - erroneously. There was no significant difference between search time and computer and search experience.

Hsieh-Yee (1998) has compared *simulated* searches for text with searches for graphic information as well as known-item and subject searches using Alta Vista. The author

suggests that due to the structure of the Web a hierarchical approach, (i.e. the searcher actively manipulates the URL of a page to access a particular level in a resource's hierarchy in order to explore it) be often used as an additional tactic to traditional tactics like keyword and author searching. The ideas put forward should be further investigated with data taken from actual web sessions.

Large, Beheshti and Moukdad (1999) investigated the moves (or actions) made by primary school pupils during web searching. Fidel and colleagues' (1999) study focused on high school students and found that they were focused and flexible searchers, but that training and search support was necessary to release the great potential of the Web as an information gathering resource. In Bilal's two articles (Bilal, 2000; 2001) on 12-14 year old children's use of the Yahooligans! Web search engine she examines their cognitive, affective, and physical behaviour when using the search engine. She has among other things compared how these use the search tool for solving tasks of different complexity and found that "children had more difficulty with the research task than with the fact-based task" (Bilal, 2001, p. 135).

2.4 Web interaction research needs

Previously performed LIS-based Web research has to a large extent been founded on methods used to study traditional IR systems. These studies to a large extent focus on characteristics of web searching that are known from studies of other search systems. Such methods cannot, in our view, sufficiently be used to reveal the interplay between the searcher and the different potential information objects that are accessible via the Web. It is therefore necessary to develop a method (schema) that takes into account the characteristics of the Web as a conglomerate of different sources as well as the characteristics of the searcher and her problem or task. We believe a description of the complex system of factors and relationships that web searching involves in the form of a *conceptual framework* is the first step needed in order to identify fruitful research problems. It proceeds towards explaining what happens during web searching and may facilitate the development of systems that support factors that are not taken into consideration today.

Such a framework could be used to shed a light on web search success factors and make it possible to focus the attention of system developers, search intermediaries, and searchers to the dimensions of user-web interaction that are open to improvement. It should have a strong theoretical basis on the one hand, which means that it should be compatible with previous well-tested attempts at describing information behaviour. On the other hand it should be applicable to real data, which means that it must be neither too general nor too

close to the data. The framework should support the organisation of data for the research problems in a fruitful way.

Although some studies exist that examine how searchers perform search tasks and, to a smaller degree, work task on the Web (Fidel et al, 1999; Large, Beheshti & Moukdad, 1999; Bilal, 2000; 2001) these do not take into consideration the direct effects of different dimensions of search tasks and work tasks on the search process.

How searchers' work tasks and search tasks influence their searching is an important area of study. But the main question we should have in mind when developing our tool is "*how is the Web used; what explains the search process*?" We are working on a methodological research question and want to identify factors that can be used to describe and analyse WIS-processes.

When people choose to use the Web in order to get help in performing some kind of task, it is necessary to identify what kind of work tasks they are involved in as well as how they specify their search tasks. Web searching should be seen as a process which involves a multitude of sources and interaction techniques and which is dynamic in the sense that it is dependent on the searcher's needs, while at the same time influencing those same needs. All these characteristics of Web behaviour may cause problems for the users and hinder their success in solving their tasks.

It is necessary to get a better understanding about how people interact with the Web in order to develop tools that facilitate web searching. Thus to formulate research problems in the context of a framework such as the one proposed above will benefit the progress of web focused information behaviour research.

2.5 The strategy of disjointed incrementalism

Braybrook and Lindblom (1963) introduced the strategy of disjointed incrementalism (DI) to explain how decision-makers treat complex problems. The strategy is very pragmatic and is characterised by decisions-makers focusing on a limited set of factors that produce small changes, while at the same time knowing that there are other important factors that are not taken into consideration. The strategy can be contrasted with rational theories of problem solving/decision-making demanding that those responsible for executing a task or solving a problem need to have total information about all factors influencing the problem and all consequences of possible solutions.

We believe the DI-strategy also is a fruitful point of departure for understanding what happens during Web information search (WIS) processes. There is a large number of factors that may influence the searcher when she interacts with the Web. To identify the relationships between all factors would be impossible. Studies of interaction with information systems where fewer factors are involved imply that searcher-system interaction is seldom characterised by rationality. We are aware of many problems that occur during information behaviour; these problems can be categorised according to the cause of the problem:

- Unclear information need. The searcher may not know precisely what her information need is. This is what Belkin, Oddy and Brooks (1982) called an anomalous state of knowledge (ASK);
- Semantic problem. The searcher experiences problems in expressing her information need, although she knows *what* she needs. Such problems may be transitory, but if the problem is complex it is often valuable for the searcher to seek help solving it, for example to explore it in a reference interview;
- Syntactic problem. The searcher experiences problems using the search system's syntax to formulate her need in the format required by the system;
- Usability problem – resource level. The searcher experiences problems understanding the presentation format of the resource;
- Usability problem – software level. The searcher experiences problems understanding how to interact with the software implementing the information system;
- Usability problem – computer level. The searcher experiences problems using the computer.

It is quite clear that searchers do not encounter search systems with perfect information about their task or the system itself. In addition they are influenced by a lot of factors that are not necessarily perfectly tuned towards the search process or indeed will *stay* perfect during the session.

When discussing the DI-strategy Braybrook and Lindblom focused on American public policy to explain how it differed from the rationalistic or "synoptic conception of problem solving" (Braybrook & Lindblom, 1963, p. 37-57). They describe eight features of decision-making that characterise the DI-strategy (Braybrook & Lindblom, 1963, p.83-106). When related to WIS-processes these features can be used to characterise four important aspects.

Restricted number of alternatives and consequences to consider

The searcher will, due to the limitation in human intellectual capacity, lack of imagination, and the multitude of alternative resources available, restrict the number of resources she wishes to explore in the process. The searcher most often will focus on those resources most familiar to her and avoid extreme alternatives; what Braybrook and Lindblom call "nonincremental changes". There may be a number of consequences that are not taken into consideration by the searcher during interaction. These may be important and the searcher may be deliberate in not taking them into account. This happens for several reasons; the searcher may find them uninteresting, remote from her problem, imponderable, intangible or she may simply not understand their importance. Searchers use their pre-understanding when deciding what resources to explore. Although they may be aware of resources that have a reputation for treating a topic of interest they still, in most cases, decide to use resources that they believe will give them a *satisfying* response.

Adjustment of objectives during interaction

During a search session the searcher may decide that it is not possible to find the necessary resources needed to deal with the search task and/or work task. There are no resources that are usable for her initial purpose and instead she will make adjustments in her task(s). In other words we can describe it as an example of the ends adjusted to the means. She may collect material that was not what she expected to look for initially in the process, because the ideal resources do not exist.

Reconstructive treatment of data

Another way of viewing the dynamics of WIS processes is to focus on the searcher's reconstruction of the search task and/or work task during the process. The results obtained in the search process constitute active components in the reformulation of the search task. For example, the searcher may learn that she had an erroneous conception of an important factor in her initial search task formulation making it necessary to adapt her new knowledge into the formulation and possibly choose a new strategy to deal with it.

Serial analysis and evaluation

The work task constituting the point of departure for WIS processes may well be executed over a long time (see also Vakkari, 2001), for example in the case of a student working on her doctoral dissertation. The point of departure for the searcher at one stage of task processing may be different from what it was months and years ago. Still the searcher may use the same or similar Web resource types at different stages in the process, perhaps using different techniques, but all the same the goal may have changed incrementally and the searcher adjusts her interaction accordingly.

For these reasons it seems natural to develop our descriptive method with a viewpoint reflecting the fact that a searcher engages in WIS-processes with a strategy which is influenced by a certain amount of knowledge and awareness of certain factors while not taking into account other relevant knowledge or factors. In our opinion the searchers are not rational individuals, meaning that they do not hold perfect information, during WIS processes. They, nevertheless, manage to interact with the system in one way or another by focusing on a few factors that at each stage seem sufficient to reach their goal. We might call them adaptive actors; emphasising their ability to adjust themselves to the conditions presented by the system. There is a need to capture those factors that the searchers seem to be aware of and those factors that can be identified as playing important roles during search behaviour in the domain of the method schema. The method schema should yield findings that are faithful to real empirical data in all its richness and adaptive incrementalism. It should not impose a rationalistic model on the actors. The domain should be in a format that permits researchers to extend and reduce the number of categories and attributes, but at the same time it should be stringent enough to be generalisable for as large a number of search sessions as possible.

2.6 The demands for developing the method schema

In order to develop our method schema we need to specify the requirements for the domain, the procedure, and the justification components.

To develop the *domain* we need to specify the categories and relationships of the problem under study. Knowledge of relevant research can be found in the literature of information behaviour (information seeking & information searching/retrieval). In addition it can be useful to perform an analysis of empirical research results on Web searching.

To develop the *procedure* we need knowledge of research methods applied to similar research problems. The literature on information behaviour may point us towards fruitful methods. In addition it is necessary to have a design for the procedure, i.e. a prescription on what sub-procedures ("methods") to apply in order to collect specific kinds of data. Since a vast number of known empirical social science "methods" may be useful, it is not possible to go into detail in discussing their application for WIS analysis. We prefer to present a strategy for their application rather than their detail and to rely on the reader's general knowledge in suitable "methods" of social science.

To obtain the *justification* of our method schema we need to specify the schema to a method that allows a strategic test of the core of the domain and the procedure. And we need to apply that method to an empirical case.

We present a number of models of information behaviour in Chapter 3 and see whether they can be used to help us in developing the method schema. In Chapter 4 we relate categories of these models along with important characteristics identified elsewhere to our domain. The models also suggest methods to be used for collecting and analysing data and are thus related to our procedure, also presented in Chapter 4.

An empirical study is needed in order to refine the framework (domain) as well as to justify the method. The framework should, at least partly, be based on existing knowledge in the field, but it will also be necessary to perform a pilot study in order to clarify what categories and dimensions as well as relationships are identifiable. The pilot study will use grounded theory techniques in order to identify characteristics of the search process that have not been described in the existing literature. According to Strauss & Corbin (1990) the use of grounded theory (GT) opens up for combinations of inductive and deductive thinking, which seems to fit our research problem and the sketched design of our study. The pilot study is also presented in Chapter 4.

3 Evaluation of previous models of information seeking and user-centred IR

In the present chapter we examine some of the most well-known models of information behaviour to see what concepts they present as important for research on information behaviour. The models have been assessed according to their *content* and *format*.

We have selected some categories from the information behaviour literature that we believe represent important dimensions of WIS behaviour. To assess the usefulness of the existing models we have looked at their coverage of these categories, which can be said to represent the content dimension.

To assess the formal requirements of the models we have used the set of criteria proposed by Bunge (1967) (see Subsection 1.3).

Our point of departure is that scientific problems do not suddenly emerge from a vacuum. It is our pre-knowledge and prejudices that necessarily make us view the world in a specific way. Our "theoretical sensitivity", i.e. the researcher's personal knowledge of his field (Strauss & Corbin, 1990), forms the perspective we use when we study a phenomenon.

It is therefore superficial to say that we look at an objective set of facts. The facts alone do not provide the necessary basis to identify the truth about the phenomenon. Cohen and Nagel (1934, p. 199) state that it "is the difficulty, or problem, which guides our search for some *order among the facts*, in terms of which the difficulty is to be removed." The identification of a problem viewed through our theoretical perspective is the foundation that makes us seek out ways to solve it.

3.1 Evaluation of models of information behaviour

Several authors (e.g. Wilson, 1981; Wilson & Walsh, 1996; Dervin, 1983; 1992; Ellis, 1989; Ellis, Cox & Hall, 1993; Kuhlthau, 1991; 1993, Ingwersen, 1992; 1996; Belkin, Marchetti & Cool, 1993; Belkin et al. 1995; Saracevic, 1996a; Spink, 1997) have developed models in which they try to capture the essence of information behaviour. Their points of departure have been partly theoretical and partly empirical, and the models' levels of abstraction vary from very general to extremely close to the information searchers' actions.

We have chosen to differentiate between models of information *seeking* and *searching* behaviour. The models have been selected based on their influence within the library and information science community; i.e. they have been heavily cited or/and are referred to in central review articles in for example ARIST – The Annual Review of Information Science and Technology (Dervin & Nilan, 1986; Hewins, 1990).

3.1.1 Evaluation criteria

Content-based evaluation criteria

We can use Engelbart's (1962) demands for conceptual frameworks to evaluate the models with respect to their content by focusing on:

1. The important factors of the model
2. The relationships between those factors
3. How changes in the factors/relationships affect the model
4. The research goals and methods that are proposed by the model

In particular we are interested in the factors and their relationships. The latter two criteria (i.e. 3 and 4) are dealt with in our evaluation of the model's ability to satisfy formal criteria. In the content-part of our evaluation we look for traces of the following broad categories (factors) that we believe play important roles when explaining WIS processes. We have put in parentheses references to literature that, at least to a certain point, treat the role of the individual categories:

- Work task (e.g. Byström & Järvelin, 1995; Vakkari, 2001; 2003)
- Searcher (e.g. Wilson, 1981; Ingwersen, 1992)
- Searcher's surroundings/environment (e.g. Hjørland & Albrechtsen, 1995; Audunson, 1999)
- Search task (e.g. Marchionini, 1995; Reid, 1999)
- The search process – including searcher-system interaction (Ingwersen, 1982; Fidel, 1985; Bates, 1989)
- Search system features (Ingwersen, 1992; 1996)

The evaluation may reveal attributes of these categories and important categories that we had not thought of. During the evaluation we should take this into consideration, i.e. all models should be examined bearing in mind the same categories and attributes. For this

reason it may be necessary to perform the evaluation in more than one round. Table 3.1 (Subsection 3.2) summarises our findings.

Formal evaluation criteria

In our review we seek to systematically examine to what degree the existing models satisfy the evaluation criteria for scientific theories put forward by Bunge (1967, p. 352-354). The models we examine are developed to map important factors of information behaviour and/or they intend to point out how to explain what happens during information behaviour. We do not claim that these models are theories or that they were ever intended to be so by the authors, but the criteria are nevertheless relevant; focusing on epistemological as well as methodological aspects. In other words we examine the models as if they were theories.

We have chosen to focus on those criteria dealing with epistemological and methodological issues and categorised them in order to see their purpose. To a certain extent we have supplemented them with demands not explicitly stated by Bunge. The criteria were introduced in Subsection 1.3, but we re-present them here with accompanying comments to explain their purpose.

It should be emphasised that this list does not represent absolute demands to scientific theories or models[8]. Rather the criteria help us determine the overall strength of the individual models, and make it possible to compare the models.

Criteria focusing on the systematisation of knowledge:

1. Compatibility with previously well-tested knowledge.
   This has to do with the model's ability to take into account what has been established through previous research as "truths" concerning the object of study.
2. Integration of formerly separate parts of knowledge.
   The model may introduce new combinations of previously established knowledge. Theories from psychology can, for example, be used to explain information behaviour or a broader context, such as work tasks, may be suggested as an addition or supplement to search tasks.
3. Generalisation and explanation of lower abstraction level knowledge (observations, data) through higher level constructs.

[8] Please refer to Subsection 1.2 in order to find definitions of central concepts used in this dissertation.

The criterion states that we should look at the level of analysis provided by the model's set of categories or concepts and it should be possible for someone familiar with the research field to easily learn how to organise this data according to the model.

4. Explanation of facts through systems of hypotheses which entail the facts.
   This criterion deals with the explanatory power of the model, i.e. it is used to examine how good the model is at proposing hypotheses for information behaviour.
5. Expanding knowledge by deducing new propositions based on selected starting points and collected information.
   The model may point out how existing facts can be used to deduce new facts.

Criteria focusing on the guidance of research:

6. Pointing out fruitful problems.
   The criterion focuses on the model's ability to propose new research problems within the field
7. Proposing the collection of data which nobody would understand to collect without the theory.
8. Proposing new lines of research.
   Criterion 6, 7, and 8 all deal with the originality of the theory. Criterion 8 focuses on its ability to guide the researcher towards the development of new research programs (Lakatos, 1970) or paradigms (Kuhn, 1996). Thus the first model that extended system-centred IR models by user-related categories proposed a new line of research.

Criteria focusing on the mapping of (parts of) reality

9. Representing or modelling the objects (and relationships) of the relevant part of reality instead of plain summarisation of data.
   The model should go beyond the surface of observations and state how significant concepts relate to one-another
10. Suggesting tools for producing new data.
    The model may include guidelines for data collection as well as analysis

We can see that the criteria suggested are not mutually independent. Criteria 3 and 9 both refer to the categorisation of the relevant parts of reality that the model is developed to treat, Criterion 9, however, relies on empirical support. Similarly ideas about new research programs (Criterion 8) would logically generate ideas about research problems (Criterion 6) and data collection.

Some of the criteria can also be seen as mutually exclusive. A new and revolutionary model, satisfying Criterion 8, may often well be characterised by not being inclusive with

respect to previous established knowledge in the field (Criterion 1). Or it may choose to treat the established knowledge as a special case.

We have taken the time factor into consideration when evaluating the models' novelty, i.e. a model proposing a new line of research in 1975 would be considered to satisfy Criterion 8 whereas a 2002 model repeating the same arguments does not.

We have not ranked the criteria and they are not supposed to be mandatory requirements of the models we examine. Their purpose is to signify the strengths and weaknesses of existing models of information behaviour, but it is quite clear that models that do not fulfil the same set of criteria can be valuable for different set of purposes.

We have decided to use a ternary score for evaluating the models' degree of compliance with the formal criteria (see Table 3.2). This is obviously a simplification; the models may satisfy criteria in different degrees – for example with respect to their ability to point out fruitful research problems. We, nevertheless, use a "+" to state that a model satisfies a criterion, a "(+)" means partial satisfaction, while a "-" means that the model does not satisfy, or only in a small degree satisfies, the criterion.

For the evaluation we have used the core documents presenting the models. We are aware that some of the models are presented in a number of documents, in such cases we have chosen to pick the document(s) most frequently cited in the LIS research literature. Table 3.2 (Subsection 3.2) summarises the individual models' ability to meet our criteria.

When applying the content and formal evaluation criteria, our focus is on the models' contribution to our research problem. Thus another application area might credit the models differently.

3.1.2 Models of information seeking behaviour

A central research area within LIS consists of defining what triggers information needs and how human beings behave when trying to solve problems, or problematic situations (Wersig, 1971), change their anomalous state of knowledge (ASK) (Belkin, Brooks & Oddy, 1982), bridge gaps in their understanding (Dervin, 1992), or execute work tasks (Byström & Järvelin, 1995).

Dervin and Nilan, in their much cited review article (1986) called for a constructivistic paradigm in studies of information need and use, i.e. an understanding of information

seekers as humans who are constantly constructing their knowledge of the world. The emphasis should be moved from viewing information systems as objective entities towards examining them from the users' point of view (p. 16).

Studies of this kind are known under many names. Information seeking, information behaviour and information need and uses studies are terms used to denote similar work. To make things even more complicated the term information seeking is sometimes used to describe information searching, which we return to later. T. D. Wilson (1999) sees information *seeking* behaviour as a part of information behaviour and information searching behaviour as a subset of information seeking behaviour.

In the LISA database approximately 350 articles have been indexed as dealing with "information seeking behaviour" since 1980. These kinds of studies have focused on e.g. users' academic background (e.g. Ellis 1989; Ellis, Cox & Hall, 1993), profession (e.g. Ginman, 1987) and task complexity (Byström & Järvelin, 1995). The object of these studies is for the researchers to try to obtain as complete as possible an overview of the users' total information behaviour. Thus the use of formal as well as informal and textual as well as non-textual information sources and channels is studied.

Wilson

Wilson's 1981 article contains two models, one for explaining information behaviour and another for describing information-seeking behaviour. Wilson claims that the information behaviour model is not an attempt "to 'model' information-seeking behaviour but to draw attention to the interrelationships among concepts used in the field" (Wilson, 1981). In this sense he suggests a meta-model and states that particularly "information use" and "information exchange" are concepts that need to be more thoroughly investigated.

His most current model of information-*seeking* behaviour (Wilson & Walsh, 1996) is a major revision of the second model from the 1981 article and focuses on how a person within a context deals with her information need. This is the model examined here; Figure 3.1 presents the 1981 model, slightly simplified after (Wilson & Walsh, 1996) while the 1996 version is summarised in Figure 3.2.

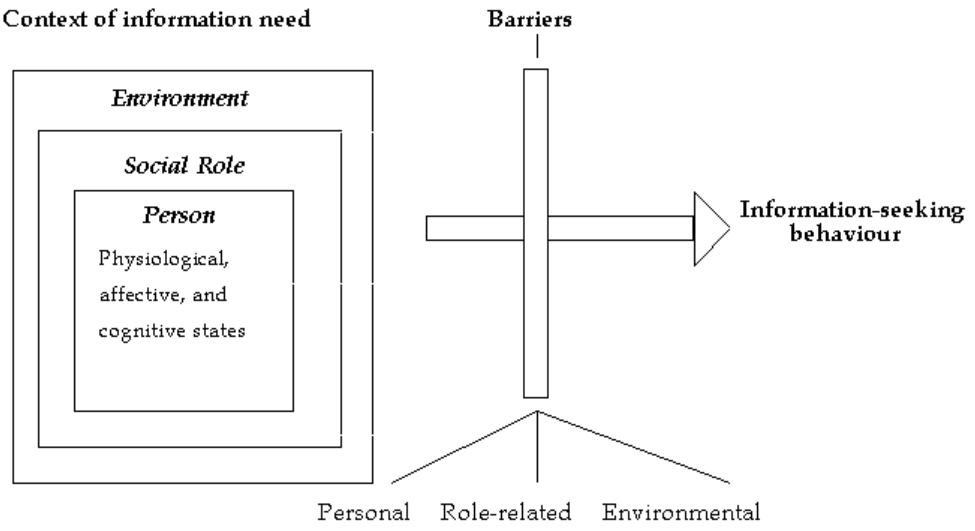

Figure 3.1 Wilson's (1981) model of information seeking behaviour

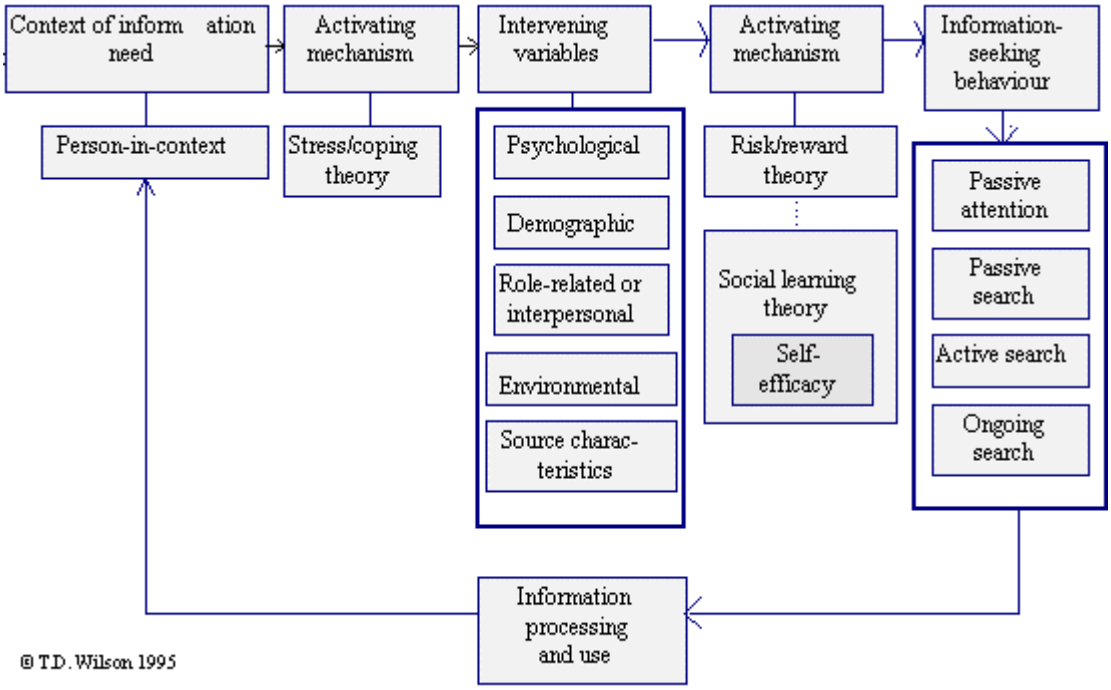

Figure 3.2 Wilson and Walsh's (1996) model of information seeking behaviour

The 1981 model introduces the context of a searcher's information needs and the barriers that may prevent him from trying to satisfy them. These barriers are examined more thoroughly in the newer model, so we will only focus on the different kinds of needs as suggested in this model. Wilson uses literature from psychology to support the division of information needs into three categories of needs that may lead to information seeking

behaviour. He states that cognitive needs will be related to a definition of information as representing "factual data" or "subject content", which, at the time of writing the 1981 paper was, and still is, probably the most common sense understanding of information. Affective needs are – according to Wilson (1981, p. 9) – more related to conceptions of information as the communication channel, for example, in the sense that one of the communicators wishes to satisfy a need for recognition or acceptance. "Physiological needs may trigger affective and/or cogntive needs" (Wilson, 1981, p. 7).

Based on thorough literature reviews of other disciplines Wilson and Walsh (1996) introduce some variables, which intervene in the seeking process. These intervening variables represent the barriers in the original model. The model suggests that personal characteristics like, for example, "selective exposure" or educational level influences the searcher's choice of information sources. It also takes into consideration how theories of "stress/coping", "risk/reward", and "social learning" at different stages may explain the continuation or discontinuation of the seeking process.

The fact that the model implicitly contains models collected from other disciplines can be seen as both a strength and a weakness. On the one hand it signifies the complexity of information seeking behaviour and indicates the ties that obviously exist between LIS and fields like psychology and organisational decision making. On the other hand the most fundamental problem with general models is that they get weak. The ambition to create a model that aims to capture the whole spectrum of information behaviour can thus be said to be an attempt to bite off too much. It is nice to have a map over a field and its adjacent areas of research, but it is probably too general to be used as a tool for developing hypotheses, although it may be used to point out fruitful problem areas and approaches.

It is tempting to call this a model of information behaviour rather than the more specific seeking behaviour. In that sense it can be seen as a revision of both models presented in Wilson's 1981 article. It should be noted that Wilson is unclear about whether this is a model of information behaviour or information seeking behaviour. He calls it the latter in his 1999 review article on models in information behaviour research (Wilson, 1999). In 1996, however, he stated that "the model can be taken to apply to information behaviour more generally, rather than to solely information seeking behaviour" (Wilson & Walsh, 1996). The original information-seeking model (Wilson, 1981) represents a simpler scenario for examining how a person with an information need is influenced by her physiological needs, affective needs, and cognitive needs when deciding to engage in information seeking. These needs are in turn influenced by her environment and the roles

she plays. We choose to evaluate the most recent model, i.e. the 1996-version, since it is clearly a development of Wilson's earlier work.

Epistemologically it seems that Wilson's model satisfies some of our criteria. It is compatible with previously tested knowledge (satisfying Criterion 1) – though on a very general level – being a summary model. Given that the purpose of the 1996 model is to summarise how knowledge derived from a variety of different disciplines can be used to explain information behaviour, it obviously can be said to be an attempt at integrating separate parts of knowledge (C2), although this is also on a general level.

Wilson should also be given tribute for being among the first to focus on the user in the seeking process, thus guiding researchers towards a new field of research and satisfying Criterion 8. Although previous research (e.g. Paisley, 1968; Allen, 1969) had focused on users' choice of information resources and attempted at modelling the user in her surroundings, Wilson (1981) was among the first to try to clarify the nature of a person's dynamic information needs.

The very general nature of Wilson's model, as well as the scarcity of guidelines for applying the model, prevents it from satisfying the other criteria.

When we look at it in the view of our research problem it becomes clear that it is much too general. It is obvious that the abstraction level of the model's units is not compatible with our needs. The model's main purpose is in our view to sketch some units that play a role during information behaviour and, since we are interested in information *searching* behaviour within a certain context (World Wide Web), there are several and more specific factors that need to be addressed.

With respect to the content-based requirements we can see that Wilson covers the searcher (person-in-context) and implicitly her environment, but there are no suggestions of attributes characterising these, in our view, important categories – thus the contribution to our work is little. Neither work task nor search task is included in the model, nor is the search process[9] or any elements of searcher-system interaction.

Dervin

Brenda Dervin has developed the sense-making methodology over many years (Dervin, 1983; 1992; 1999), and it could have been represented before Wilson's work since it " has

[9] The *seeking* process is modelled in Wilson's work, but that is on too general a level for our purposes.

been under development since 1972" (Dervin, 2001). She calls it "a set of metatheoretic assumptions and propositions about the nature of information, the nature of human use of information, and the nature of human communication" (Dervin, 1992, p. 61-62). The sense-making approach uses a metaphor to describe how people solve problems. One is asked to assume that a person that faces a problem – a gap - will construct "a new or changed sense" – or build a bridge over this gap with the use of some kind of help (p. 68) - see figure 3.3.

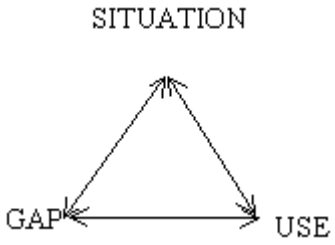

Figure 3.3 Dervin's (1992) situation-gap-use metaphor

The original purpose of her methodology was to develop techniques to identify how information needs occur and are solved. A process perspective is advocated where the focus is on situations where problem solvers face problems, how they are able to solve them, and how they continue working afterwards. The strategy of "neutral questioning" (Dervin & Dewdney, 1986) is developed in order to help researchers formulate questions that can reveal what characterises such situations. The methodology can be criticised for its strong emphasis on the "situation", which makes it extremely focused on the individual and the problem(s) she faces in a given situation. It may provide good help for librarians and information intermediaries who face a person with an information need. In our view the "model" leaves out too many issues that influence information searching and is probably more useful as a tool, in Dervin's words, "concerned primarily with theorizing of the metatheoretical sort" (Dervin, 1999, p 729). When testing the model[10] against our criteria we focus on its usability as a tool for analysing information needs and seeking, not its metatheoretical functions.

Dervin's metaphor is much narrower than Wilson's model in the sense that its point of departure is one specific way of dealing with information needs. More important it should be treated as a collection of methods focusing on how to do research on information

[10] Note that Dervin does not call the sense-making approach a model, but among other things a "set of methods" (1992, 61).

behaviour (and related areas). Dervin's model is a very general representation of user-problem interaction, so general that it can be said to be compatible with previous knowledge (C1). We do not think it satisfies any of the other criteria focusing on knowledge systematisation. The metaphor can be used to point out research questions (C6), and it certainly advocated a new line of research on information behaviour at the time, thus satisfying C7, which has to do with suggesting new forms of data to collect, as well as C8; "proposing new lines of research". The neutral questioning technique could also be considered as a proposal for *how* to collect data (C10). However, it cannot be said to be a model made for representing reality in the sense of identifying objects and relationships.

It is also clear that the model only in a minor degree satisfy the criteria generated by our problem, i.e. our content-based criteria. Its focus on the process of information behaviour is, however, relevant and it is seen from the perspective of an information seeker (person) facing gaps (information needs) in her understanding, thus it is fair to say that a certain conceptualisation of the search process can be identified.

For our purposes it is, nevertheless, hard to see how Dervin's approach could be used since we want to understand WIS processes generated by work tasks and search tasks, and the sense-making methodology suggests nothing about how to analyse work and search tasks. Neither does Dervin's work address Web information searching and the features that are specific when it comes to interacting with different resource types and search techniques.

Ellis

David Ellis (1989; Ellis, Cox & Hall, 1993[11]) has identified eight categories that can be used to describe individuals' information seeking behaviour:

- Starting: activities characteristic of the initial search for information;
- Chaining: following chains of citations or other forms of referential connection between material;
- Browsing: semi-directed searching in an area of potential interest;
- Differentiating: using differences between sources as filters on the nature and quality of the material examined;

[11] In Ellis and Haugan (1997) they identify "identical or very similar features of information seeking behaviour". We have chosen to use the original categories in our examination.

- Monitoring: maintaining awareness of developments in a field through the monitoring of particular sources;
- Extracting: systematically working through a particular source to locate material of interest;
- Verifying: activities associated with checking the accuracy of information;
- Ending: activities characteristic of information seeking at the end of a topic or project, for example, during the preparation of papers for publication.

The researchers identified these eight categories by interviewing different academic user groups (social scientists and natural scientists) about their information seeking. The interview data were analysed with the help of grounded theory based procedures (Glaser & Strauss, 1967). In Ellis (1989) they examine how the categories can be supported by IR systems and specifically propose the use of the hypertext format for development of systems that may facilitate such seeking strategies.

The categories are not necessarily carried out in any specific order although such an order is suggested by the presentation above. Ellis (1989) claims that the interrelationships of the categories are person-situation specific and therefore can "only be indicated in the most abstract and general terms" (p. 178). Ellis, Cox and Hall (1993) states that "the models[12] do not attempt to define the interactions and interrelationships between the categories or the order in which they are carried out" (p 359).

Although the categories focus on information seeking it may be that they can also be used to explain information searching by categorising occurrences of particular patterns of user-system interaction as being, for example, monitoring or extracting. This is particularly interesting when studying user-Web interaction since it involves such a variety of sources within one "system".

The researchers point out the fact that many studies of information seeking behaviour differ in "aims, objectives and methods" (Ellis, Cox & Hall, 1993, p. 356) making it impossible to compare them. The authors state that their research is an attempt at rectifying these defects and that they hope to be able to make a "method of investigation in user studies which can be employed to obtain comparable results across a range of different subject groups".

[12] The plural s refers to individual "models" for each of the user groups studied. Both groups of scientists "undertake similar activities and the sources employed are also similar" (Ellis, Cox & Hall, 1993, p. 365).

Previous studies constitute "background information" (Ellis 1989, p. 172). Due to the lack of well-tested knowledge within information seeking at the time we cannot see that it satisfies C1. The categories can, however, be seen as generalisations or abstractions of lower level knowledge, thus satisfying C3.

The model was at its publishing date original with respect to its focus on seekers' behaviour and as such satisfies the 8th criterion. Also its use of grounded theory guide researchers towards a tool for data production and analysis, thus also satisfying C10.

In Subsection 1.2 we have chosen to define a model very loosely and therefore cannot claim that Ellis' work does not satisfy our definition. On the other hand, we cannot claim that it is a representation of the entire problem area that it explores, particularly because it does not state the relationships between the categories. Whether the model can be characterised as a source of hypotheses (C4), we are not sure; Ellis states that "starting may lead to chaining, differentiating may play a role in identifying sources" etc. (1989, p. 179). Hypotheses should be stated more strictly. Wilson (1999, p. 256) points out a way to treat the categories as steps in information seeking processes, thus explicating the relationships more. We have indicated this in Table 3.2 in Subsection 3.2 by setting the score in parentheses. This discussion indicates, however, that it can be identified as a source of potential research problems (C6), although its major contribution has been as a tool for classifying information behaviour.

In view of our research problem we can see that Ellis' work does not handle Web searching specifically. The search categories can be used to describe a dimension of the work task, which we may call the work task process. There are no dimensions that one may relate to the search task. Thus there is no support for research on how aspects concerning the search task affect the search process. Neither are there any guidelines on how to order process data useably and identify their attributes. There are indications of a seeking process in his model (what we have called the work task process), but it is not explicitly stated how this affects searcher-system interaction.

One may be able to identify all Ellis' categories in WIS processes. On the other hand it is also possible to see a Web session as dealing with one single category – indicating a stage in the searcher's work task. In this sense the model may serve as a source for ideas on explaining Web searching, for example by using the categories Ellis suggests to broadly categorise Web search behaviour.

Kuhlthau

Carol Kuhlthau (1991; 1993) has developed a model of the information seeking process (ISP, unfortunately she has called it the information *search* process) which takes into account the information seekers' emotional, cognitive and physical experiences at different stages of the process. The six stages in the ISP and their accompanying tasks are

1. Initiation – recognise information need
2. Selection – identify general topic
3. Exploration – investigate information on general topic
4. Formulation – formulate focus
5. Collection – gather information pertaining to focus
6. Presentation – complete information search

In Figure 3.4 the relationships between actions, and emotional and cognitive experiences in the individual phases are explicated.

Stages	**Initiation**	**Selection**	**Exploration**	**Formulation**	**Collection**	**Presentation**
Feelings	Uncerta-inty	Optimism	Confusion, frustration, and doubt	clarity	Sense of direction/ confidence	Relief/satis-faction or disappoint-ment
Thoughts	vague	----------	----------	----------	---------->	focused
Actions	Seeking relevant informa-tion	----------	----------	---------->	Seeking pertinent information	

Figure 3.4 Kuhlthau's (1991) model of the information search process

These stages in the seeking process are normally entered chronologically by information seekers who work with large problems. Typically the seeker experiences uncertainty in the first stage and moves towards relief in the last. Smaller routine problems may be solved without the need of information seeking at all, or the stages may pass within a moment.

As stated above the similarities of Kuhlthau's and Ellis' works have been explicated by Wilson in a recent review article (1999) by addressing the categories of Ellis to specific stages of Kuhlthau's model. In our opinion this is the way to highlight the virtues of both researchers' work. The problem with Kuhlthau's model being its general nature and lack of specification with respect to how the tasks are performed functionally, this can be a bit

more emphasised by applying Ellis' categories to the stages. Her focus on information seeking as a *process* does, on the other hand, direct the focus towards an important characteristic of information behaviour on both the macro as well as the micro level. The main purpose of the ISP model is to guide librarians so that they can better understand and meet users' needs at different stages of their information seeking.

Kuhlthau (1993) claims that the model is based on the ideas of constructivism as advocated by Kelly (1963), Dervin's (1983) sense-making, Taylor's (1968) levels of information needs as well as the cognitive viewpoint in information science (e.g. Belkin, 1990) – thus it can be said to satisfy the integration criterion (C2).

Like Ellis' work Kuhlthau's model is grounded on empirical data and it contains a generalisation of factors that systematically appear in her data (C3). She has performed several longitudinal studies inspired by her experiences as a school librarian. The model states the relationships between the different stages in the seeking process as well as the interrelationships between the cognitive, affective, and physical actions that may take place in the individual stages, although this is on a very general level. On the other hand it does not take explicitly into consideration important dimensions like searcher, work task, and search task. It can, nevertheless, be said to satisfy the criteria of modelling a relevant part of the world rather than purely being a summarisation (C9). She also documents methods that can be used to collect relevant data, thus Criterion 10 is also satisfied. Kuhlthau's work has inspired many research projects; in pointing out research problems (C6); by providing ideas about the kind of data to collect (C7); and it may well be argued that she represents a new line of research in information behaviour (C8). It can further be argued that the model suggests hypotheses by stating relationships between physical, emotional, and cognitive actions which can be tested in other circumstances (C4). One might, for example, focus on different user groups, which is what Kuhlthau herself has done (Kuhlthau, 1999; Kuhlthau & Tama, 2001).

Seen in the context of our needs we observe that it does not specifically address Web searching. The unit of analysis is projects that last longer than single Web sessions. Since the model assumes that searchers are involved in complex work tasks, its stages can be used to emphasise one significant dimension of work task performance. We do not, however, believe that Kuhlthau's model alone is sufficiently usable for describing WIS processes for several reasons:

- It does not explicitly model work tasks;
- It does not focus on other possible work task attributes other than "stage";

- It does not take into consideration individual search tasks or other dimensions that may influence the process;
- The searcher is only an implicit dimension in the seeking process;
- It treats information seeking/searching on a too general level.

3.1.3 Models of information searching behaviour

Information *searching* is used to denote activities that relate to the interaction between a person with an information need and an information system. The term "information system" is more comprehensive than "IR system". It can for example be used to describe different kinds of web resources. The models that are presented in this section all are developed to describe information searching, which means that the *searcher-system interaction* is the focus of the models. In addition we will see that the models to a different degree also incorporate knowledge about factors that influence this interaction. In particular, many of the researchers have been interested in the searchers' mental or *cognitive* processes during information searching.

Ingwersen

Peter Ingwersen (1992; 1996) has developed a cognitive model for IR interaction based on analytic and empirical user-oriented studies. The model addresses how the different participants dealing with the information retrieval system (such as the searchers, the indexers, and the system designers) are influenced by their cognitive spaces, as well as by their social and organisational environments (see Figure 3.5). It can be seen as a general model for understanding all actors taking part in IR.

A central factor in this model is that of the knowledge structures that are embedded in the different actors in IR interaction. Through our interaction with the surrounding world we exchange experiences with others which in turn may lead to changes in our knowledge structures. A person's knowledge structure thus constitutes his internal model of the world (Wersig, 1971, p 57-58). But, as referred to in the discussion about the cognitive viewpoint in Chapter 1, Ingwersen (1996) also takes into account the fact that the "mental activities take place in the context of a domain(s) of epistemic, social or organisational nature" (p 10).

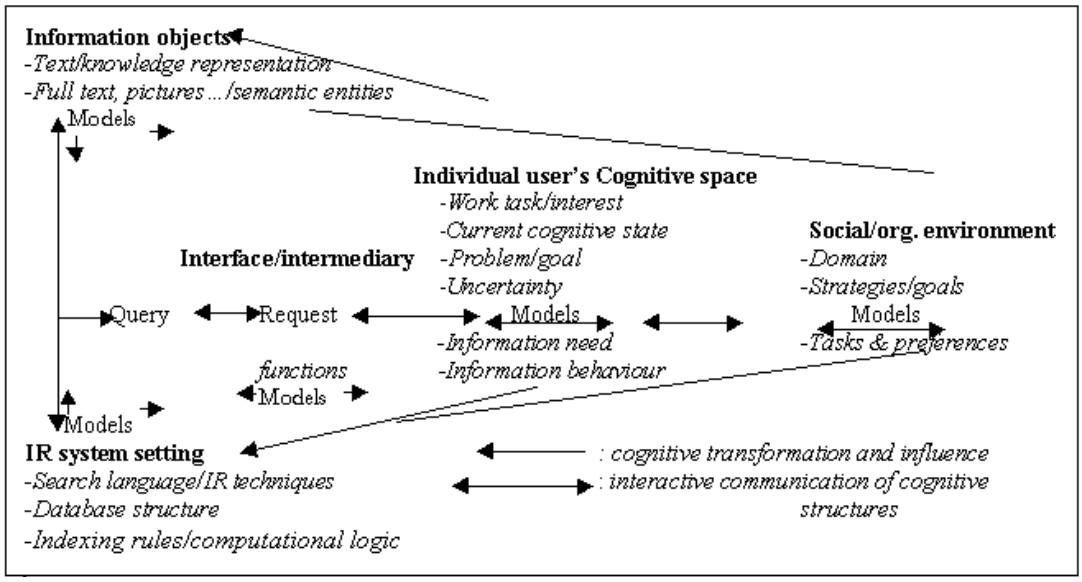

Figure 3.5 Ingwersen's (1996) cognitive interactive information retrieval model

There is no doubt that this model is a representation of issues that are important for information searching, although it lacks an explicit process view. The model's strength is its attempt at integrating issues related to information seeking and IR, it is categorised as a model of information searching due to its focus on the searcher that interacts with the IR system. The criterion dealing with integration of knowledge (C2) is thus satisfied and so is the generalisation criterion (C3). In our view it does not satisfy C4 (explanation of facts through system of hypotheses) because no such hypotheses are explicitly stated. On the other hand the model is a valuable source for developing research problems and hypotheses. It does, however, satisfy C5 by developing ideas from previous works representing the "cognitive viewpoint" (e.g. Belkin, Oddy & Brooks, 1982).

From a methodological point of view we can see that the cognitive IR interaction model seems to satisfy all criteria (C6, C7, and C8). It has clearly been a strong advocate for a new line of research within IR, being the most thorough explication of the cognitive viewpoint and being quoted as background knowledge in many papers.

On the other hand the cognitive viewpoint has been criticised (e.g. Frohmann, 1992) for being too focused on cognitive aspects and paying less attention to, for example, emotional factors. Cognitive changes may also be hard to spot since it, at least in its purest form, requires that the searcher constantly explains her thinking during the interaction with the information system.

If we assess the model with the help of our formal criteria, many of them seem to be satisfied. That does not necessarily mean that it can be applied to Web information search processes. We will briefly discuss how it can be used before focusing on our content-based criteria.

The model opens up for many hypotheses about the effects of the relationships between the objects, but it is questionable whether its main use is to investigate user-system interaction. Empirical work based on the model has been done to develop a method for evaluating interactive IR systems (Borlund & Ingwersen, 1997; Borlund, 2000). Ingwersen (1996) advocates the use of "polyrepresentation", i.e. to incorporate representations of all participating cognitive structures into the user-system interaction and he convincingly argues for use of more realistic settings when testing search systems.

Ingwersen's model can be characterised as holistic, i.e. it can be seen as an attempt to describe the main objects that constitute the field of library and information science (LIS). Research on librarians; their attitudes, background, education etc. may, for example, be seen in the perspective of this model's *intermediary*. The *IR system setting* covers IR systems design, and text analysis can be viewed in the light of the *information objects*. When combining the model's objects we will see that the user – information – social/organisational environment triangle is covered by e.g. literature sociology. In this sense Frohmann (1992, p. 371) may be right in pointing out that the cognitive viewpoint offers a "total theory for LIS".

Returning to our Web searching related criteria we may see that Ingwersen's work is applicable to WIS processes in the sense that it recognises process related searching – "IR can be thought of as a discontinuous, non-linear process". The process view is, however, not explicated in the model and there are no accompanying guidelines on how to incorporate it. The model does not suggest attributes within the process which can be related to external factors. It nevertheless lists factors that play important roles in the different cognitive structures, including the searcher and her social/organisational environment. These factors are not examined in detail and due to the model's viewpoint the categories and possible attributes are to a large degree designated as parts of the searcher's cognitive space. Factors like "work task", "problem/goal", "uncertainty", "information need", and "information behaviour" are presented in the model as attributes of the user's (searcher's) cognitive space, the environment has "domains", "strategies", and "tasks", and the information objects (system) has a content part as well as a system setting (structure). Ingwersen's cognitive model of IR interaction also is the only model

that we have evaluated that includes IR system features, such as search language, indexing rules, and database structure.

In his book, Information retrieval interaction, Ingwersen (1992) introduces a second model – the Mediator model, which we deliberately have not included in this evaluation. Mediator can be seen as a successor to the Monstrat model (Belkin et al, 1987). These models are developed to design interactive IR systems and take into consideration a number of functions that need to be understood by the system's intermediary mechanism. The functions discussed by Ingwersen (1992, p. 204-220) cover many important properties of the searcher as well as the IR system. Due to both models' (Monstrat and Mediator) focus on the intermediary mechanisms and functions of the IR system, we have chosen to omit them from our analysis. However, Ingwersen's 1996-model, covers many of the Mediator model's non-system factors which seem relevant to use in explaining information behaviour.

Ingwersen's model would need to be applied to a far more detailed extent if we wanted to look at the *context* in which the search *process* takes place, in other words external factors that affect the system-user interaction. The model would need to be elaborated and specified in order to analyse the effects of, for example, work task and search task on WIS processes. Belkin's episodic model, which is presented below, is in this sense closer to the actual interaction.

Belkin

N. J Belkin and his colleagues (Belkin, Marchetti & Cool, 1993; Belkin et al., 1995) sees IR interaction as consisting of IR episodes, a series of which constitutes a search process, and suggests a model for classifying searchers' information seeking (or in our terminology *searching*) strategies, or ISSes. The model has been used not only to create interactive interfaces for existing IR systems, but also to develop a prototype system – MERIT (Belkin et al., 1995). An ISS is described by its position in the four dimensions: method (scanning or searching); goal (learning or selecting); mode (recognising or specifying); and resource (information or meta-information). The models' dimensions are binary, thus making it possible to describe 16 different, but idealised, kinds of ISSes (see Figure 3.6).

It is not clear why these particular dimensions were selected and the authors themselves claim that the model should be subject to empirical research.

ISS	*Method*		*Goal*		*Mode*		*Resource*	
	Sc	S	L	S	R	S	I	M
1	X		X		X		X	
2	X		X		X			X
3	X		X			X	X	
4	X		X			X		X
5	X			X	X		X	
6	X			X	X			X
7	X			X		X	X	
8	X			X		X		X
9		X	X		X		X	
10		X	X		X			X
11		X	X			X	X	
12		X	X			X		X
13		X		X	X		X	
14		X		X	X			X
15		X		X		X	X	
16		X		X		X		X

Method: Sc=Scan; S=Search Goal: L=Learn; S=Select Mode: R=Recognize; S=Specify
Resource: I=Information; M=Meta-information

Figure 3.6 Belkin's information seeking strategies (Belkin, Marchetti & Cool, 1993)

We consider the ISSs to be more of a classification system for interaction episodes than a model representing a research problem. There are no indications of relationships between units in the scheme, although the authors point out there might exist dependencies, "especially in the relationship between the dimensions of *method* of interaction and *mode* of retrieval" (Belkin, Marchetti & Cool, 1993, p. 326) – we therefore cannot see that the strategies go beyond summarising the data. If we look at it in the light of the criteria dealing with systematisation of knowledge, the authors claim that it accounts for certain features of seeking behaviour found in, among others, Ellis' 1989 model. But the authors make no attempt at positioning the seeking strategies in the landscape of previous attempts to describe information behaviour. The dimensions obviously generalise search behaviour (C3), but, perhaps due to its foundation within the cognitive viewpoint, the model lacks important factors that depend on dimensions other than the mental ones.

The classification system may be characterised as a methodologically fruitful one (C6). Given that we have an IR system that we wish to test, the classification can be used to investigate which search strategies the system or, for example, web site has problems in meeting. On the other hand the dichotomic values of the ISSes may be too abstract to state the essence of what actually happens during searching.

It is easy to agree that information searching can be treated as episodes, but an important issue is how to define the episode. The ISSes are probably on too low a level to be related to work task and search task; the goal being too strategy-specific. We cannot see that either task type or the searcher is treated in the model. The low level makes it difficult to explain the interplay of external factors and thus make sense of the process. With modification one may use the ISSes to analyse WIS strategies, or WISSes (Pharo, 1999), but the value of results generated from such analysis may not be worth the effort. The dimensions are mainly related to the process/interaction-category. In addition "resource" can be said to represent an aspect of the information system. We do not find that the information seeking strategies meet any of the other content-related criteria. The conceptualisation of searching into episodes – at some level – is, however, relevant and should be taken into consideration when describing WIS processes.

Saracevic

Tefko Saracevic (1996a) compares the problems involved in IR interaction with those expressed by researchers on human computer interaction (HCI). His stratified model (Figure 3.7) describes users within a situation and an environment that interact with computational IR systems at three different levels: the surface level; the cognitive level; and the situational level.

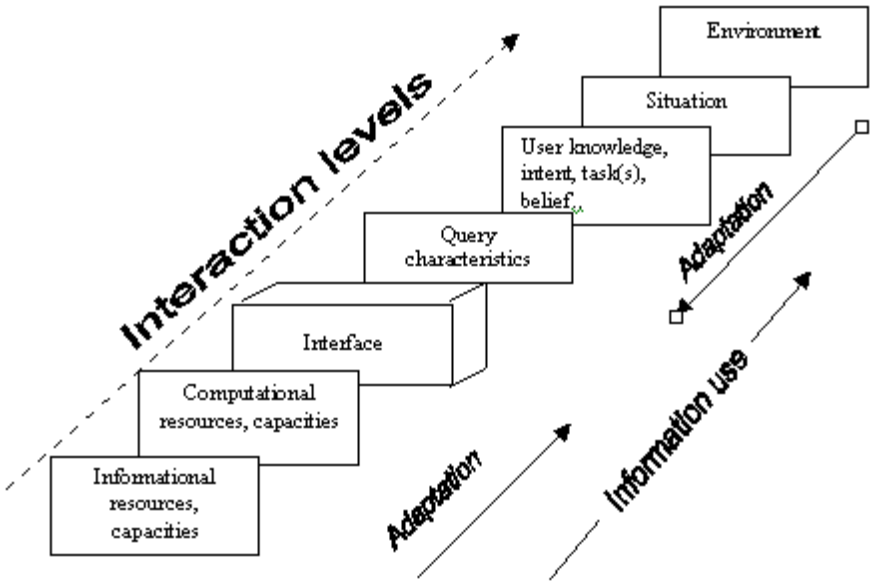

Figure 3.7 Saracevic's (1996a) stratified model of IR interaction

The levels, which each contain several elements, can be treated separately, but as Saracevic points out (1996a, p. 7), changes that occur on the deeper situational and cognitive levels influence the actions that take place on the surface. For example the query term selection, and the query field used, will be a result of the documents or lack of documents retrieved by the previous query formulation. If we look at it from a hypertext perspective, for example a searcher using the World Wide Web, the potential information found may make the searcher change the resource (type) used as well as the interaction method, i.e. she may follow links rather than entering queries.

Like most of the models presented here, Saracevic's model is a theoretical construction and, as far as this author can see, not directly based on empirical investigations. Nevertheless there is little doubt that both his and Ingwersen's models, which are quite similar, are particularly good at describing the dynamic area of IR interaction, with the focus being on a user in a (problematic) situation. Unfortunately Saracevic, like Ingwersen, only implicitly treats the interaction as a process – he states that "each strata/level involves different elements and /or specific processes" without it being specified in Figure 3.7, which only "lists the elements involved in the levels during interaction" (Saracevic, 1996a, p. 6).

The model lacks indications about how its factors influence one another. Although traces of a process-oriented view can be found in the arrows linking the levels, there are, however, no guidelines on how to apply it neither for the collection nor the analysis of the data in order to understand the process. The factors themselves seem to be on an appropriate level of abstraction (C3) and the model is compatible with the other cognitivly oriented models of information behaviour, i.e. Ingwersen's (1996) and Belkin, Marchetti and Cool's (1993). Although it is arguable whether these models should be considered to be "well tested". It has more explanatory power than the ISS model since it involves more factors as well as factor-attributes, for example searcher and situational characteristics. In addition it includes search system features. One may, perhaps, use the model to generate hypotheses concerning the relationships between more factors/attributes, but Saracevic has made no such hypotheses explicit.

The model is fruitful with respect to generation of research problems (C6). No guidelines about its use, i.e. about what data to collect and how to analyse them, are present in the article.

Concerning WIS processes, overall, Saracevic's model provides some good ideas about categories and attributes that play important roles in Web searching, but it suffers from its

inability to model how the process influences the various factors. Her "knowledge", "tasks", and "beliefs" influence the searcher. She works within a context (situation) and an environment to solve her problem/task. The model is quite similar to Ingwersen's and it can be seen as an operationalisation of the latter, focusing on the user-system interaction. It can be criticised for being too holistic to be usable for actual analysis. We believe that it suggests relevant factors that influence WIS processes, but that it has severe weaknesses that hinder its direct application to our problem – being too holistic, having no direct process-orientation, and too little analytic power with respect to the relationships between categories.

Spink

Spink (1997) presents a model of the interactive search process based on an empirical study and developed using grounded theory. Her study primarily focuses on the role of *feedback* in interactive IR. She states that "as a central dynamic of the interactive IR process, feedback involves the user's evaluation of the IR system output, user's judgement, and query modifications". She goes on, stating it is an area that "is intriguing and not well understood in IR, with different views of feedback in the research literature" (Spink, 1997, p. 382). Five categories of feedback are identified; content relevance feedback; term relevance feedback; magnitude feedback; tactical review feedback; and terminology review feedback. Her work on this important problem is interesting and we would refer the reader to her text. For our purpose we focus on the model she uses to describe the interactive search process, which we believe should be treated as an extension or operationalisation of the interactive IR model (Ingwersen, 1996).

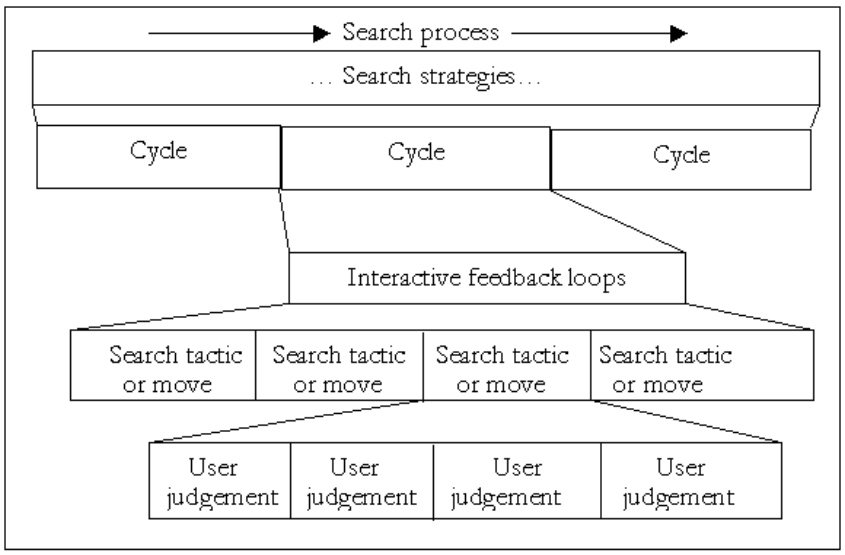

Figure 3.8 Spink's (1997) model of the interactive search process

Spink's model (Figure 3.8) deconstructs the interactive search process into finer levels of granularity, beginning with the *search strategies* that consists of *cycles*, which in turn are made up of *interactive feedback loops*. The feedback loops may consist of *moves* or *tactics*, and the user's interpretations or judgements accompany his/her actions. We may well consider this a generalisation of lower abstraction level knowledge, meeting our third criterion (C3). In addition her work proposes hypotheses about the nature of different forms of interaction feedback (C4).

The model does not specify how external factors influence at the different levels in the search process. The focus on feedback is interesting and the phenomena that generate the different feedback categories should be examined further in order to see if they are relevant for inclusion in a model dealing with WIS processes. For example, it may be that the searcher's choice of actions is dependent on the magnitude of resources that she faces (related to magnitude feedback (Spink, 1997, p. 388)). The model takes into consideration and develops previous work on feedback (C5 is satisfied), but when evaluating its compatibility with established knowledge we should focus on its claim that it extends the interactive model of IR – and it appears to do so. It clearly points out fruitful problems (C6) to investigate and exemplifies data analysis (C10).

The model focuses on traditional IR systems with the use of a homogenous database, not the Web environment. It does not explicitly relate to work task, search task or any of their potential attributes. A similar critique can be made against Spink's model as against Belkin's ISSes; the individual moves may be on too low a level for identification of interplay.

Since it was developed first and foremost as a tool to study feedback in IR, it is unfair to demand that it should be suitable for other purposes. In combination with more general theories (such as the ones presented above) it may be used to identify interplay between features of information seeking and interactive IR.

3.2 A discussion of the existing information behaviour models

The fact that so much effort has been put into model building signifies a scientific school that is trying to establish its field of study (Kuhn, 1996). Before a field of study has established a scientific paradigm, i.e. reached a common understanding of its main problems and methods, there will be competition between scientific schools that wish to achieve hegemony within the field.

The works examined above do not differ considerably in their points of departure, which may mean that a user-centred paradigm is beginning to become set within the field of "information behaviour". The focus on individual users' behaviour has been criticised, as exemplified above, and the social constructivistic view has to a certain degree been incorporated in, for example, Ingwersen's model. It should also be noted that the field constitutes a small social environment and given that its core members have reached an "agreement" about the focus of study, it might be difficult for "outsiders" to be heard.

The models have been evaluated in order to see how they satisfy our criteria (Table 3.1).

	Wilson	Dervin	Ellis	Kuhlthau	Ingwersen	Belkin	Saracevic	Spink
Work task	-	-	-	-	(+)	-	(+)	-
• Stage	-	-	+	+	-	-	-	-
Search task	-	-	-	-	(+)	-	-	-
• Information need	+	(+)	-	-	(+)	-	-	-
• Strategies	-	-	-	-	-	-	-	+
Searcher	+	-	-	-	+	-	+	-
• Uncertainty	-	-	-	+	+	-	-	-
• Work task	-	-	-	-	+	-	+	-
• Problem/goal/ intent	-	-	-	-	+	-	+	-
• Information need	-	+	-	-	+	-	-	-
• Model/knowledge	-	-	-	-	(+)	-	+	-
Environment	+	-	-	-	+	-	+	-
• Domain	-	-	-	-	+	-	-	-
• Strategies/goals	-	-	-	-	+	-	-	-
• Tasks	-	-	-	-	+	-	-	-
• Preferences	-	-	-	-	+	-	-	-
Search process/ interaction	-	-	-	-	+	+	+	+
• Method/query/ move	-	-	-	-	+	+	+	+
• Goal	-	-	-	-	-	+	-	-
• Mode	-	-	-	-	-	+	-	-
• Resource	-	-	-	-	-	+	-	-
Search system	-	-	-	-	+	-	+	-
• Search language	-	-	-	-	+	-	-	-
• Indexing rules	-	-	-	-	+	-	-	-
• Database structure	-	-	-	-	+	-	-	-

Table 3.1 Summary of existing models' ability to satisfy our content-based criteria

We summarise our findings below and start by looking at the criteria focusing on content. In Table 3.1 we present a criteria-model matrix where "+" indicates explicit support of

category or attribute, "-" means that the category or attribute is not supported by the model while "(+)" means that the model includes the characteristic, but treats it as an attribute of another category.

We can observe some similarities between the models, although only Wilson, Ingwersen, and Saracevic explicitly include the *searcher* in their models, she is the driving force behind the activities in the other authors' works. The 'problem' or 'information need' is another central concept, which in some cases has been treated as an attribute of the searcher.

Search system features are only treated by Ingwersen and, at a very general level, Saracevic. Neither of these two models explicates how or whether, for example, work task relates to system features.

Only Ellis and Kuhlthau can be said to focus explicitly on the processing of tasks, but neither have clearly defined the object that generates information searching. Kuhlthau uses the term "task" ambiguously, sometimes it seems to be related to the problem or work task at hand (Kuhlthau, 1993, p. 38), and at other times the different stages in the process represent tasks (Kuhlthau, 1993, p. 120). Ellis refers to "activities [searchers] engaged in which might be understood as having an information component" (Ellis, 1989, p. 174). Both Ellis and Kuhlthau have as their main object of study stages that information seekers need to go through during problem solving, indicating that *stage* is an important attribute of work task. Ingwersen and Saracevic, on the other hand, treat work task as an attribute of the user. The other models do not suggest work task as a factor influencing information behaviour.

Search task is not explicitly mentioned as a category in any of the works we have examined. The concept of "information need" is in our view a dimension of search task, since in a sense it constitutes the goal of the search task. Information need is a central category in Wilson's summary model and is treated as a user-attribute by Ingwersen. Dervin (1983) claims that the gaps are translated "in most studies as "information needs"", but we choose to classify her way of treating information need as an attribute of the user (searcher). Spink does not explicitly mention search task, but "search strategies", which is the point of departure in her model, clearly deal with how to solve search tasks.

"Searcher", "person", and "user" are the terms that have been used in the examined models to denote the person in need of information. Ingwersen and Saracevic have both equipped the searcher with a series of attributes, some of which, we believe, should be placed elsewhere. In our view it would be suitable to let the attributes of the searcher be

those that are of a general nature and let dimensions like "problem" and/or "task" be treated separately with their own attributes.

The "context" in Wilson's "person-in-context" is the role and environment in which the person experiences the information need. Unfortunately Wilson does not make any suggestions about how it can affect the search process nor does he examine any of the features of the environment any closer. Ingwersen suggests, on the other hand, many attributes that would be interesting to examine as influencing WIS processes.

The four authors presenting models of information searching are the ones that appear to be closest to the actual interaction; they all suggest "method" as an interaction-attribute. In addition, Belkin's ISSes include three other dimensions that all deal with aspects of interaction. It is clear that neither Wilson, Dervin, or Kuhlthau take into consideration the interaction during information searching. Kuhlthau states that the "actual actions taken" (Kuhlthau 1991, p. 362) should be one of three aspects of a model of information seeking. In her model she does, however, focus on the remaining two parts, the affective and the cognitive aspects rather than how the searcher "physically" acts as a result of emotional input or changes in her cognitive model. Ellis is closer to the interaction since at least some of his categories are applicable to system interaction, for example browsing and monitoring.

In Table 3.2 we summarise our assessments of the formal criteria used on the models. "+" means that the model satisfies the criteria, "(+)" means that the criterion is partly satisfied, while "-" indicates that the model does not satisfy the criterion. A full list of all criteria can be found in Subsection 3.1.1.

	Wilson	Dervin	Ellis	Kuhlthau	Ingwersen	Belkin	Saracevic	Spink
C1	(+)	(+)	-	-	-	-	-	-
C2	+	-	-	+	+	-	-	-
C3	-	-	+	+	+	+	+	+
C4	-	-	(+)	+	-	-	-	+
C5	-	-	-	-	+	-	-	+
C6	-	+	+	+	+	+	+	+
C7	-	+	-	+	+	-	-	-
C8	+	+	+	+	+	-	-	-
C9	-	-	-	+	-	-	-	-
C10	-	+	+	+	-	-	-	+

Table 3.2 Summary of the models' ability to meet formal criteria (list of criteria in Subsection 3.1.1)

None of the models satisfy all formal criteria, although Kuhlthau's model is very close. Please consult the sections on the individual models for more thorough discussions. We see that Criterion 1 (C1), which deals with the compatibility of previously well tested knowledge, is not completely satisfied by any model. Wilson (1981) and Dervin (1983) whose models represent early attempts at modelling user-oriented information behaviour, on such a general level that it can be considered indisputable, is credited by a partial match. This is due to the general lack of "well-tested knowledge" in the field and most models can be characterised as pioneer works within the area of information behaviour. Some of the authors, e.g. Kuhlthau, can be seen as operationalisations of well-established ideas and Belkin, Saracevic, and Spink are all related to the cognitive model of IR interaction. The lack of empirical support in previous work has, however, made most models fail Criterion 1.

Wilson in his 1996-model, Kuhlthau, and Ingwersen, on their respective levels, try to integrate previously non-integrated knowledge (C2) to their models. For our purposes Ingwersen is the most interesting since his model is an attempt at combining issues of information searching (retrieval) and seeking.

Six of the models map the world on a level that seems applicable for the categorisation of empirical data. Both Wilson's and Dervin's models are too general for that purpose and thus do not satisfy C3 as such.

Only Kuhlthau, Spink, and, at least implicitly, Ellis satisfy the fourth criterion, which has to do with the model's ability to explain phenomena in the form of hypotheses. The remaining models (minus Wilson's) may well serve as sources for hypotheses and point out research problems (C6).

Only two of the models that we have evaluated (Ingwersen's and Spink's) have been found to satisfy C5, i.e. expanding knowledge as a result of known facts. This may be due to the pre-paradigmatic nature of the LIS field. The models are often grounded in the authors' own empirical work and thus they generate their propositions inductively.

C7 deals with proposing the collection of new forms of data and is satisfied by both Dervin and Kuhlthau. Wilson, Dervin, Ellis, Kuhlthau, and Ingwersen[13] additionally can be seen as pointing towards a new user-centred paradigm, satisfying C8 as well.

[13] Belkin also belongs in this group, his ASK hypothesis (Belkin, Oddy & Brooks, 1982) focused on characteristics of searchers' information needs.

The last two criteria (C9 and C10) deal with the mapping of reality. C9 specifies that the model should be more than a pure summarisation of data, which we believe is satisfied only by Kuhlthau of the seeking models. The four models that are grounded on empirical data or have been empirically tested document the data collection and analysis methods and can therefore be said to satisfy C10.

Saracevic's and Ingwersen's models are quite similar, partly because Saracevic's model to a certain extent is built on Ingwersen's. They take into consideration the dynamics of searcher-system interaction and specify factors which clearly are the driving forces behind the dynamics. The models clearly are usable as sources of research problems, organisation of data collection, and categorisation of results. Their lack of explicit process orientation necessitates that we must go beyond them and really investigate the search process.

Both Kuhlthau's, Spink's and Ellis' models are developed using grounded theory (Strauss & Corbin, 1990), which means that they have been inductively derived from the data. Belkin's information seeking strategies were tested against data collected on a group of students using the Web (Pharo, 1999), leading to the conclusion that it was necessary to omit the mode dimension as well as to modify the existing dimensions to make the strategies suitable for that purpose (see Chapter 4).

Kuhlthau focuses on activities related to information behaviour and bases her research on longitudinal user studies. Her model is however very general, which means that the actual interaction between user and information system is not treated in any detail in her work. It would therefore not be applicable to our use.

A similar argument can also be used against Wilson's model. It is difficult to relate it to actual use of information systems. Both Wilson's and Kuhlthau's models do however open up for formulating hypotheses about how specific intervening variables, or stages in the search process, influence the use of a particular search system.

Dervin's situation – gap – use metaphor is a tool that could be used in developing expert systems, i.e. one might provide a system with a feature that was able to relate specific situations with identifiable gaps to specific "uses". To be suitable for analysing search sessions it is however necessary to classify situations into more refined levels.

Ingwersen's model satisfies more criteria than any of the other information search models that we have assessed in this study. This suggests that it would be practical to pursue when developing a framework for explaining WIS processes. As stated above, however,

we should keep in mind that the interactive IR model has been developed for other purposes than "purely" describing search behaviour and that the latter is probably not its best virtue.

None of the existing models can be used as methods for solving our problem, but they provide categories and attributes that can be applied in a conceptual framework to describe characteristics of WIS processes. Further categories and attributes can be found in the literature on LIS and can be inductively derived from data on WIS processes.

4 Designing the method – problem statement and procedure

This chapter presents the method we have used in order to design our method schema[14] for analysing Web interaction, in other words our meta-method. We first provide some theoretical background focusing on the treatment of complex problems, which we believe justifies the strategy we have chosen to deal with our research problem. Then we summarise the path we have used to develop the schema in Subsection 4.2, before we – in Subsection 4.3 – go on to examine the initial sketch of our domain.

In Subsection 4.4 we go through the grounded theory-methodology and show how we have used it to develop a conceptual framework for studying WIS processes. The framework is described in the fifth section of this chapter. In Subsection 4.6 we discuss the procedure of our method, i.e. how to use the framework. In the final section of this chapter (4.7) we show how our framework relates to the existing models of information seeking and searching.

4.1 Ill-structured problems need flexible methods

Newell (1969) has called those problems "ill-structured" for which no formalised method of solution has been developed due to the problems' fuzziness or complexity. Computers or formal algorithms cannot solve ill-structured problems. These need to be explored by human beings in order to be organised in a solvable form. Only human problem solvers possess methods flexible enough to solve problems above a certain level of complexity.

In the preceding chapters we have shown that web information search processes are dependent on a lot of features related to the person(s) involved in the search process, his/her environment, the work task executed by this person which leads to the WIS process, the search task directly generating the WIS process, and, last but not least, features of the search system and the search process itself. There may be other important categories that we have not identified and there are certainly attributes of the categories presented that we have not taken into consideration. We can, however, easily imagine that to get a complete picture of the WIS process with all intervening variables would require enormous effort in data collection and analysis.

[14] For convenience we use the term method in this chapter, but the reader should be aware of the distinction. When we find it strictly necessary to pinpoint the particular features of method schemas we use this term.

Methods can be evaluated with respect to their generality and their power. We will discuss the implications these properties of methods raise when the problem is "ill-structured", as is the case with our problem.

A problem of this complexity signifies that a multitude of factors play significant roles in its solution. It is necessary to define the most important of these factors and identify the relationships between them. In addition the method needs to address techniques that can be used in order to identify "the values" taken by the factors.

The more factors that are added to a method the more complex it will be. It may therefore be fruitful to generalise it to a method schema (Eloranta, 1979), which we will discuss later.

On the other hand, as the number of problem types the method schema will be applicable to rises so does the method's generality. One measure of a method's fruitfulness is to see how many different kinds of problems it applies to, and thus we may compare it with other methods that are developed to address problems in the same domain.

Unfortunately there is an inverse relationship between the generality of a problem and its power; when the generality increases the power weakens. With (Newell, 1969, p. 372) we define the power of a method to be a measure of its ability to deliver solutions. Three dimensions of the power of a method can be identified:

The *probability of solution* refers to the method's ability to solve problems within a domain. The more problems it is able to solve the more powerful it is.

The *quality of solution* must be measured in terms of the proposed solutions. If two methods produce solutions that differ in level of explanation we may say that the one with higher explanatory power, if this is the only difference between the methods, is the most powerful.

The *consumption of resources* is the third dimension of power. This is used to measure the investments needed by the method in order to solve a problem. Such a measure will often be time, but Newell also points out financial expenses and computational resources as possible units of measurements.

Our aim must be to develop a method schema that is general enough to cover the domain of task-based information searching in the World Wide Web. Given that Web searching is of the complexity that our examination suggests, it may also be that such a schema could cover information searching in other systems. On the other hand it must be stressed that it should be powerful enough to suggest and test significant hypotheses about WIS processes – we do want to be able to use it to focus on observable characteristics of interaction between system and problem solver.

Due to the complexity of WIS processes we aim to develop a method *schema* instead of a plain method. There are a large number of factors that influence Web information searching. The method schema should propose a *strategy* for researchers performing studies of WIS processes. This means pointing out factors that play a role in "causing" the problem as well as strategies for data collection and analysis. In the schema we should introduce categories that represent important properties of the WIS process as well as categories that suggest external factors influencing the WIS process. Thus the factors left unspecified by the schema should be easy to specify into the suitable category.

The formal criteria of Bunge (1967) (see Subsection 3.1.1) are useful as guidelines for the development of the method schema. Very roughly one could claim that the first five criteria relate to the domain, the research guidance criteria (Criteria 6-8) focus on procedure elements, while Criterion 9 and 10 can be used to point out whether the method works, i.e. if it can be justified.

An alternative to our schema approach would have been to develop a general method for analysing all kinds of task-based WIS processes, or even worse all kinds of WIS processes. Such a method would run the risk of becoming very general. Instead we make a schema that makes it possible to generate methods applicable for particular cases of WIS processes, and this method can then be examined in the light of the dimensions stated above.

4.2 The development of the method

The process of creating our method schema has gone through many stages before reaching its current form. We start the chapter by presenting our method, i.e. the method for creating our method or, in other words our meta-method. We have illustrated the research process in Figure 4.1, having numbered the boxes and indicated relationships with the help of arrows to help the reader understand the successive order of the process.

In the text we refer to the numbers in the boxes as well as to the subsection where the reader may find the result of the particular part of the process.

The presentation may look very structured and surveyable, and it is of course an idealised version of the research process. In reality we jumped forward and backwards between these stages during the process, thus there should be far more arrows in the picture to make it realistic. The use of grounded theory, which has been important in the development of our framework, i.e. the description of our *domain*, advocates interchanging between data scrutinising and category building, which makes it hard to identify exactly when the different versions of the domain were constructed.

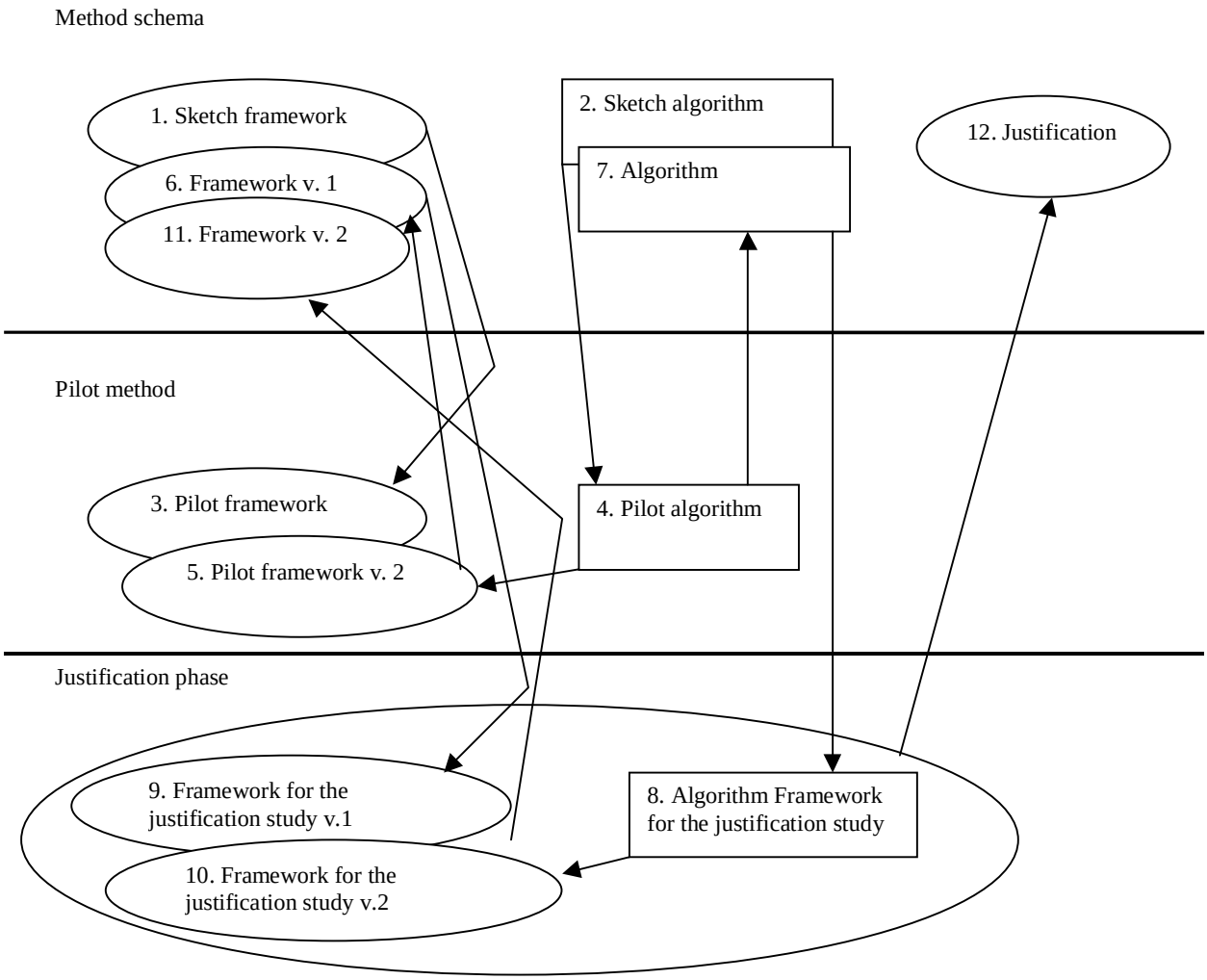

Figure 4.1 The research process

The literature on web searching reviewed in Chapter 2 and the review of existing models of information behaviour in Chapter 3 reveal a set of factors that are thought to play a role during WIS processes. We ended up with some major categories that we believe

cover important aspects of web searching, as well as information behaviour generally. The categories deal with the searcher, her organisational environment, the work task, the search task, the search system, and the search process. In addition we have looked at information behaviour literature that does not specifically present new models, but that proposes important dimensions, or attributes, of these main categories. The grouping of these findings into categories with attributes and relationships formed our first attempt at describing web information search behaviour, making it a preliminary sketch of our conceptual framework (1) (Subsection 4.3.1).

Reading the literature on information behaviour as well as literature on general methods for data collection and analysis suggested a set of "algorithms", i.e. the *procedure* of our method, that was thought to be suitable for collecting and analysing data on web information searching (2) (Subsection 4.3.2). Using a subset of the possible data collection and analysis "algorithms" (4) – and a pre-understanding about the kind of categories of WIS behaviour we should collect information on (the searcher, environment, work task etc. constituting a preliminary version of our pilot framework (3)) – we collected our data on web searching (see Subsection 4.4.1). The data we collected should be used for two purposes:

First in order to conduct a pilot study to identify additional factors (categories and attributes) to include in our framework. This pilot study (Subsection 4.4) was performed according to the guidelines of grounded theory, which suggest procedures for interpreting the data, and it returned a revised version of our pilot framework (5) in turn generating the first version of our method schema's general conceptual framework (6). To sum up we now had developed a conceptual framework representing our domain (Subsection 4.5) and we had a set of algorithms (4, refined in 7) to be used for collecting and analysing the data, constituting our method schema's procedure (Subsection 4.6).

To justify our method (12) we made our *second* attack on the data, bringing with us the framework (9) and algorithms (8), in order to reveal what they could tell us about the influences of the different categories and attributes on the WIS processes. We developed a set of empirical research questions. Next the whole data set was analysed using grounded theory, making it necessary to do some revisions of our framework (version 2 (10)), which also reflects the general framework (11). Having reclassified all data with the framework we analysed the research questions by doing a "statistical" analysis of our WIS-process stories (part of 8), and examining representative examples of the results (Chapter 5).

The justification of the method (steps 8 to 12) will be examined in Chapter 5. We will in the present chapter focus on the development of our domain and procedure.

4.3 The sketch of the schema

4.3.1 The sketch of a framework

The complexity of the problem determines the complexity of the method that is needed in order to examine it. To get a good understanding of Web information search (WIS) behaviour we believe it is necessary to use a method that takes into consideration a multitude of resource types and interaction techniques as well as other relevant properties. The fact that a human being is the active part of WIS behaviour of course makes the problem complex and properties of WIS behaviour should thus reflect the dynamics and reflectivity that characterise human behaviour.

In order to attack the problem we developed a conceptual framework as discussed in section 3.1.1, specifying the factors, relationships, research goals, and "methods", i.e. procedures, relevant for our 'system', i.e. the object of study. The 'system' in our case consists of a series of interaction episodes[15], within a searcher's context. Thus we should focus on the properties and/or objects that we believe play a significant role during WIS behaviour. Although the framework by definition also covers the *procedure* of the method, we choose to separate our discussion of domain elements and procedure elements.

We had as our point of departure made a sketch of a framework (domain) based on literature studies (see Chapter 2 and 3). This sketch pointed out a number of features that we believe play important roles during WIS processes. We believed, supported by the literature, that it would be possible to identify relationships between our information seeking-related categories (work task, search task, searcher, and searcher's environment) and the WIS process. The first attempt at conceptualising the problem is represented in Figure 4.2

Five main categories were present at this early stage. As can be seen in Figure 4.2 the categories had been developed to different levels of detail. We do not discuss the individual attributes here, as they will be discussed and supplemented by other attributes later.

[15] The "episodes" could be related to Belkin, Marchetti and Cool (1993) conceptualisation of information searching, but on a slightly higher level.

The attributes characterising "search situation" were not fully developed at this stage in the research process; they represented ideas based on studies of the information behaviour literature.

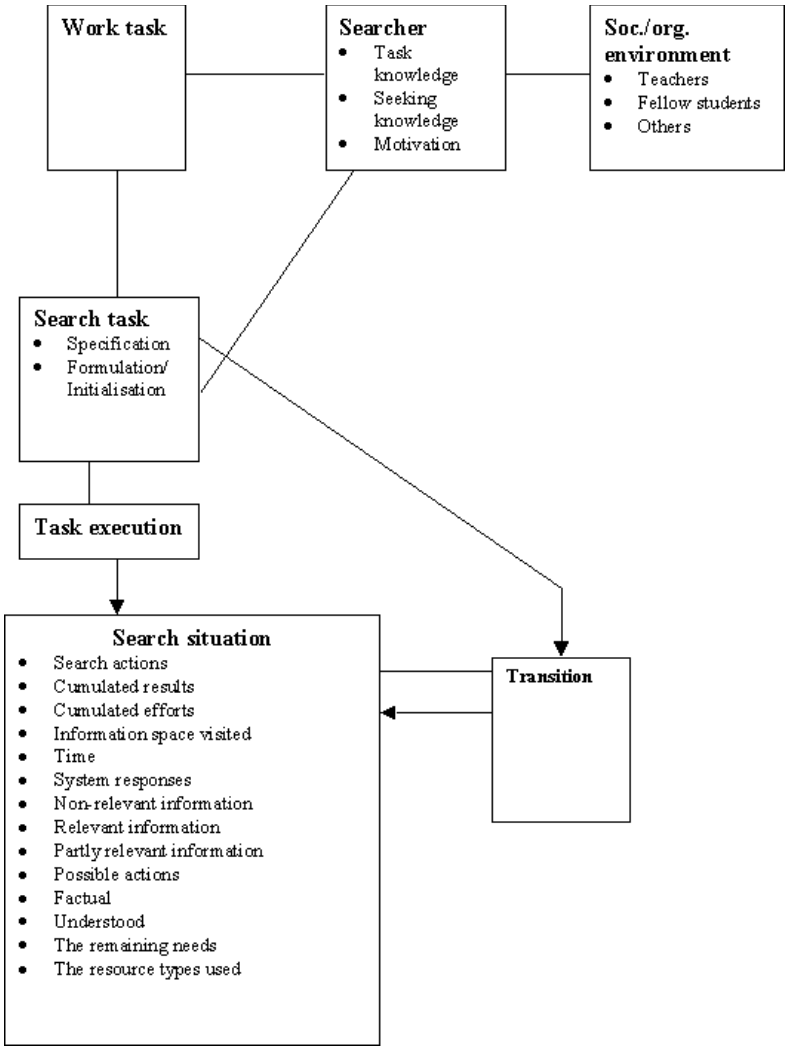

Figure 4.2 An early sketch of the domain

The *work task* represents the object that generates the information need leading to a *search task*, which in turn is executed in the form of a series of search *situations* and *transitions* (the search process). The situation and transition concepts can be seen as an elaboration of Belkin's resource type (Belkin, Marchetti & Cool, 1993; Belkin et al., 1995), which was classified as being either "information" (situation) or "meta-information" (transition). It was further thought that factors related to the *searcher* and her *environment* in some way, directly or indirectly, were related to the situations and transitions, but that this needed further investigations in order to identify what relationships actually exist.

We initially thought that the information search strategies (Belkin, Marchetti & Cool, 1993) could be used to describe the WIS-process and we later show how it was applied and tested on empirical data. When deciding what kind of data to collect in order to test the ISSs, we did, however, want to obtain as rich a material on WIS processes as possible. Therefore we considered a broad spectrum of collection methods ("algorithms") to include in the procedure.

4.3.2 Elements of the procedure

General literature on research methods proposes a variety of different ways to collect data on human behaviour; qualitative approaches (Denzin & Lincoln, 1994) include the use of in-depth interviews, observation, and textual data, for example diaries/journals. Fidel (1993) presents an overview of qualitative research in IR.

The most common quantitative approach is to use questionnaires in large scale surveys (Hellevik, 1999). Generally the goal of qualitative research is to collect a rich set of data from few respondents, while quantitative research aims at collecting relatively few variables from a large number of respondents.

We hope that the method schema can be used to generate methods requiring either kind of data, but believe it is foremost applicable as a tool for qualitative research.

Previous studies of information behaviour exemplify the use of both kinds of approaches, but focus has been on qualitative research in general since the "paradigm shift" in the late 70s/early 80s. According to Dervin and Nilan this change of paradigm was characterised by the focus shifting from an "objective" view of the user towards seeing users "as beings who are constantly constructing [information]" (Dervin & Nilan, 1986, p. 16).

Ellis (1989) – interviews, and Spink (1997) – videotaped interviews and observations, both used qualitative data collection methods in order to develop their models. Kuhlthau (1993) used various data collection methods, both qualitative and quantitative; she used qualitative methods in her first experiments and shifted to a quantitative approach in her verification studies. Some of Kuhlthau's qualitative methods were various forms of written statements made by the informants, interviews, and questionnaires. To verify her model she used a specially constructed survey consisting of 6 questions designed to "elicit cognitive and affective aspects of the information search process" (Kuhlthau, 1993, p. 94).

Several studies of search behaviour have been designed using quantitative methods. Logging user requests to information systems is one way of collecting data which can be used to analyse user-system interaction. It is an easy way of getting large amounts of data to analyse statistically, but it can also be used for qualitatively focused analysis. Such logs can be recorded on the server or client side, which is an important distinction when dealing with web interaction. Client-side recording makes it possible to track searchers' behaviour over different servers (Catledge & Pitkow, 1995; Wang & Tenopir, 1998; Wang, Hawk & Tenopir, 2000) thus seeing how the process evolves. Server-side recording, on the other hand, opens up for evaluation of interaction on that particular server by a multitude of searchers. The surveys made by Silverstein and colleagues (1999) and Jansen, Spink, and Saracevic (2000) both used query transaction logs to analyse searchers' queries in two different search engines. The demographic surveys made by GVU (GVU's WWW user surveys, 2001) were conducted using on-line questionnaires. Lazinger, Bar-Ilan, and Peritz (1997) have used questionnaires to examine Internet use by faculty members in various disciplines.

There are few studies on web information search behaviour using qualitative methods of data collection. We have therefore supplemented our review with studies of search behaviour in other hypertext systems (Marchionini, Lin & Dwiggins, 1990; Rada & Murphy, 1992; Leventhal et al. 1994). Fidel and her colleagues (1999) use observation and interviews in examining the web searching behaviour of high school students. Wang and her colleagues (Wang & Tenopir, 1998; Wang, Hawk & Tenopir, 2000) used a combination of pre-search questionnaires containing, among other things, standard tests on cognitive and affective style, screen capturing accompanied by recordings of the searchers' thinking-aloud, and post-search questionnaires.

Marchionini, Lin and Dwiggins (1990) combined questionnaires, keystroke/mouseclick-recordings, and simulated recall to study the effects of search and subject expertise in hypertext information searching. Rada and Murphy (1992) compared different hypertext systems by simply measuring the time elapsed from when the searcher started solving a given search task until it was finished. Leventhal and others (1994) used video recording and self-evaluation tasks to record how two user groups (children and adults) used a hypertext library catalogue to solve given search tasks.

Previous studies show that a variety of "methods" can be used to collect data on human information behaviour. The problem one wants to examine will strongly influence the choice of data collection methods in the specific method's procedure. If one aims at investigating how a small group of users use the Web when solving similar work tasks it

will be practicable to use in-depth interviews and observation. If the goal is to learn about how people with a particular education use the Web to deal with one particular kind of work task then it may be a good idea to use questionnaires combined with client transaction logs – and, for example, present simulated work task situations to the searchers (Borlund, 2000).

4.4 The pilot study

The aim of the pilot study was to identify relevant categories and attributes to include in the conceptual framework (domain).

We chose our collection methods based on what the general research literature suggested, but in particular we examined the methods used in research on hypertext and Web search behaviour and found that we would be justified in combining a variety of methods in order to get as rich a data set as possible.

4.4.1 Data collection

The data used in our project were collected over a period of five months from third year students in library and information science (LIS) at Oslo University College. A mandatory part of a librarian's education in Norway is to write a thesis, which is credited as 1/4th of a school year or 5 credits. All data collected were related to the students' work with these theses. The data consist of three different types of material: results of a questionnaire; interviews; and observations. In addition the complete theses have been used to gather supplementary data.

The most important type of data that were collected was centred on Web search sessions performed by a group of students. This resulted in 9 observed and video-recorded search sessions supplemented by pre- and post-search interviews. The data collected for the project were used in two ways. Three sessions were used for the pilot study. For the justification of the method these three were supplemented by the remaining six sessions to perform a study in order to analyse a set of empirical research questions. We will discuss the implications of re-using the same session for both method development and justification in Subsection 5.3.

Questionnaire

A questionnaire was sent to all third year students at the LIS department (Appendix 1). They were asked to record details about their experience in using computers and different Internet applications (Web, news, email), the topic of their thesis, at what stage in the problem solving process they were, and their current and planned use of different search sources. The students were also asked to volunteer for observations of their thesis-related web sessions.

In all 53 students returned the complete questionnaire after having been reminded three times, making the percentage of response 48 %. Of all students, 14 volunteered to be observed. This is a low response rate and lots of effort was invested in trying to recruit more students; e.g. class meetings, posters in class rooms and other significant locations, and e-mail reminders. Due to the resource limitations available for the project we decided to continue using the respondents who were available.

Interviews

Before each search session we conducted an interview with the participating student(s) (Appendix 2). The students were asked questions about the progress of their project, the sources they had used, their goals for the upcoming session, specific search strategies et cetera. The interview was designed to be open-ended, the highest priority being to have the students tell as much as possible about their intentions and plans. When two students performed the search session together they were also interviewed together.

The pre-session interviews were different in length, lasting from only a few minutes to a quarter of an hour.

After each session the searchers were asked questions about its success, what alternative sources they could have used, and how they planned to continue their work.

Observations and session logging

The students were instructed to contact the researcher whenever they wanted to use the Web as part of their thesis work. To make the Web sessions as natural as possible the students used their own accounts to log into the local network and use their own version of the client software (Netscape 3.03) with their own bookmark files etc. The users were also allowed to work together if they wished to do so. The sessions were recorded in an undisturbed room and with the help of a GrandArt video converter, which is a device that converts computer screen signals to video format. A microphone is connected to the

converter and the users were asked to talk aloud (Ericsson & Simon, 1996) during the sessions, in order to make it easier to identify their intentions and goals. Their utterances were recorded simultaneously on the videotape.

Nine Web sessions, in total 12 1/2 hours of data, were analysed. Of the nine sessions three were performed by two students working together.

Theses

The final theses were also used in our analysis. The sessions generated three theses, one of which was supplemented with a Web resource. In addition we have used syllabuses from the Faculty of Journalism, Library and Information Science and talked to Faculty staff to learn about the demands to theses at this level.

4.4.2 Introducing grounded theory

We have used a grounded theory based approach to develop the conceptual framework. Glaser & Strauss (1967) developed grounded theory and the methodology has been used in many scientific disciplines, among those library and information science (Ellis, 1989; Spink, 1997) as indicated above. There are many books on the subject and we have used Strauss and Corbin (1990) as our basic reference work. The reader should note that by "approach" we indicate that the methodology is not necessarily rigorously followed in every aspect. In particular we have used more deductive methods than we believe is usual in GT. We hope to give as complete an examination as possible of the paths we have followed to develop our framework.

More specifically, our use of grounded theory has been to apply to our analysis of the data the coding procedures advocated by grounded theory methodology and specified in Strauss and Corbin's book Basics of Qualitative Research (Strauss & Corbin, 1990). We do not claim that we have generated a theory, but our domain description is strongly related to a theory since it is in the form of a framework of categories and attributes with relationships. The relationships just remain abstract, i.e. they are not qualified or quantified, which will be the task of empirical studies applying the method schema.

Grounded theory (GT) is a qualitative method that uses a combination of induction and deduction to develop a theory about a phenomenon. Strauss and Corbin emphasise the inductive elements of GT, but in our view it is quite possible to combine deduction and induction. The researcher's theoretical sensitivity, which is a fundamental concept for

grounded theorists, forms his way of studying a phenomenon and is also the point of departure for hypotheses developed independent of his data.

GT is characterised by developing theories that are close to the data, the researcher developing the theory through back and forth interplay with the data. He develops concepts through coding the data in several stages at the same time as his hypotheses, stemming from his theoretical sensitivity, are verified by testing them on the data. Although we develop a conceptual framework, defined as being broader than theories, the methodology appeared suitable for our purposes. We had a rich set of empirical data about WIS processes that we wanted to systematise and conceptualise, using our background knowledge on information behaviour, information seeking, information searching, and IR.

Systematic analysis, or "coding", of the data will reveal its properties and relationships. The coding may be divided into three different stages, but often different kinds of coding may take place simultaneously.

The first stage (the so-called "open coding") involves categorising and classifying properties of the data and the identification of significant categories/concepts and their attributes/properties. "Axial coding" is the term used to denote our construction of the individual categories and attributes and our attempts at identifying connections between the elements. In the final coding stage ("selective coding") one identifies the core category and how it is related to the remaining categories. Below we examine the stages used to build our conceptual framework with the help of GT, but first we present the empirical data used in our pilot study.

4.4.3 Grounded theory applied on web search sessions

Three sessions (sessions 1, 4, and 9) were selected for use in the pilot study. These three were selected because students who worked on three different theses (work tasks) performed them, and the sessions represented different kinds of search tasks. Thus they represent a variety of work tasks which in turn generated different kinds of search tasks. We wanted as broad as possible a data set in order to ground a general framework and then use the remaining sessions to test the framework.

The sessions were studied in three different formats: the real time sessions, the transcribed video recordings, and the stories created from the transcriptions. We will

explain below how the different versions of the sessions were treated and how they relate to the coding stages in grounded theory.

In addition the pre- and post-search interviews, and the questionnaires returned by the users participating in these sessions were analysed and used to supplement the sessions. The interviews and questionnaires were particularly useful when doing the story analysis.

4.4.4 Real time data analysis

During the searchers' interaction with the Web the researcher sat in the background observing the process. What appeared to be significant events during the interaction were recorded on paper. These notes have been used as supplemental data for the transcription of the video recordings. Some extracts from these notes are reproduced below. The notes were originally in Norwegian, but for this purpose they have been translated into English with the Norwegian original on the right hand side. Text in italics refers to utterances by the searcher (note that only those utterances written down at the time are referred), text in parentheses refers to actions, text in brackets is used to explain utterances, while the remaining text refers to observations made by the researcher.

The main purpose of the analysis of the observer's notes was to identify major properties of Web interaction and, if possible, relationships between those. As we shall comment on later, not all findings struck us with lightning force. Some properties or attributes of Web interaction were anticipated, and thus "verified" by the data. In grounded theory terminology we can call this the open coding; we observe the data in order to group similar categories.

Example 1 – a searcher interacting with the newspaper "Dagbladet"

In Example 1 the (single) searcher is looking for information about the children's (and mystery) writer Roald Dahl. At this point of time the searcher has looked for, and found, information about the author in his Norwegian publisher's web resource and she is now looking for book reviews in Norwegian newspapers. She started by entering an index to book reviews in all papers, but this did not contain any entries on Roald Dahl. She therefore decided to inspect one newspaper at a time. The first newspaper she entered was "Stavanger Aftenblad", where she was unable to find any relevant entries. From there she went to Dagbladet, the second largest tabloid paper in Norway. To access these papers' book review indices the searcher used a meta-index. The notes below were made during the searcher's examination of Dagbladet.

Dagbladet – literature	Dagbladet –litteratur
Book reviews	Bokanmeldelser
(scans a bit)	(blar litt)
(enters query in Dagbladet: Dahl, Roald)	søker i Dagbladet: Dahl, Roald
this [= the hit list] *I don't really understand*	*dette* [= treffliste] *skjønner jeg ikke helt asså*
long list that contains entries on both roald and Dahl	lang liste m/treff på både roald og Dahl
(selects an entry tagged "reviews")	går inn på liste merket "anmeldelser"
(reformulates query dahl+roald)	reformulerer søk dahl+roald
perhaps meaningless	kanskje meningsløs
no this won't work	nei dette fungerer ikke

Table 4.1 Excerpt from notes made during a searcher's interaction with the newspaper "Dagbladet"

Example 2 – two searchers interacting with the subject index Kvasir

The next example is collected from a session were two searchers look for information about the Norwegian author and translator Jo Tenfjord. This extract is from the start of the session, and the searchers have decided to use the Norwegian subject index "Kvasir" to look for information about Tenfjord.

Meta-resources	Metaressurser
(Kvasir query tenfjord, jo)	(Kvasir søk tenfjord, jo)
17 hits	17 treff
Hits on only "jo"	Treff på "jo" alene
Norwegian translators union	Norsk oversetterforening
(Browses and examines their web to find her name)	(Browser og leter gjennom veven deres for å finne navnet hennes)
Finds it on the original page "the Bastian award" as a receiver of the award	Finner det under den opprinnelige siden "Bastianprisen" som prismottaker
Commercial for "the great feast and festival handbook"	Reklame for " den store fest og høytidsboka"
The Library Central: commercial/order form for "authors portaits"	Biblioteksentralen: reklame/bestillingsskjema for "Forfatterportretter"
Self-registered persons at the University of Bergen	Selvregistrerte personer ved Universitetet i Bergen
Strange that this should contain anything	*Rart at det skulle være noe der*
(interrupts search)	(Avbryter søk)

Table 4.2 Excerpt from two searchers' interaction with the subject index Kvasir

Example 3 – two searcher interacting with the subject index Yahoo

The third example reproduces two searchers who look for information about usability-related issues and the Web. They select the subject index Yahoo for this purpose, having spent some time using the search tool Infoseek for the same purpose.

Try Yahoo	Prøver Yahoo
(Browses "Computers and Internet")	(Browser "Computers and Internet")
Looks technical	*Ser teknisk ut*
(Browses upwards, selects "Internet")	(Browser oppover, velger "internet")
(Select World Wide Web)	(Velger World Wide Web)
(Query "usability")	(Søker på "usability")
58 hits – *too many*	58 treff – *altfor mye*
(Continues to an automatically generated query in Hotbot)	(Fortsetter til automatisk søk i Hotbot)
(Browse the first hits)	(Browser gjennom de første treffene)

Table 4.3 Excerpt from two searchers' interaction with the subject index Yahoo

4.4.5 Real time analysis results

We can see from the examples that some potentially important characteristics of Web search behaviour (or the search *process*) are identified already at this stage. In particular we may note that a variety of interaction techniques were identified. All of the examples show searchers who vary their interaction techniques between entering queries, browsing or scanning pages, and selecting links that for some reason trigger their interest. The examples also show searchers' reactions to the search results as well as to other features presented by the system.

Below we present the major results from the examination of all notes made during the three sessions.

The three session-notes were thoroughly examined and each observation was identified and categorised. Some dimensions of the search process quickly emerged and it was decided as unnecessary to categorise every occurrence of specific types of web pages (e.g. "search engine", or "commercial"), or names that represent a specific type of web page (e.g. "Yahoo" or "Alex"[16]) or specific interaction techniques (e.g. "scans page", or "enters query"). It should however be mentioned that whenever something appeared that could be said to reveal new properties of web page types or interaction techniques it was included in our categorisation, in order to get as rich as possible picture of the dimensions. These two dimensions – "resource type" and "action" - it should be mentioned, were very much expected to be found – thus they are not inductively derived

[16] Alex is the name of an author index made by the large Norwegian newspaper Aftenposten. It is regarded as an authority of its kind

from the data as such. We will return to the formal definition of the individual dimensions (categories) later and will then also relate them to findings in the literature.

The analysis of our notes resulted in the identification of six, possibly seven, dimensions of web search processes that seem to be of particular importance. In addition there are hints of other properties to look for in the more detailed levels of analysis. We will now look at these dimensions, which we have decided to name resource type, action, need, results, relevance judgement, and information space.

The *resource type* represents the type of web pages that the searchers interact with. Many different types of web resources exist and there are many ways to classify them. One way may be to separate between meta-resources and those that contain "real information". Other classification schemes could be developed using, for example, size, publisher, or function/intention as the point of departure. It is also possible to use genres developed in the non-digital world to denote the resource types, into for example newspapers, journals, and monographs. Some examples of resource types identified in the notes were "egenlaget ressursside" (Home made (meta) resource page), "Kvasir" (Norwegian subject index), "Bok & Samfunn" (Norwegian literary journal), "British humanities index" (database), "The Gate" (American e-newspaper), "WWW virtual library", "forfatterliste" (list of authors), "Gyldendal" (Norwegian publisher), "Alta Vista", and "side om Roald Dahl" (page about Roald Dahl).

Actions are used to describe the techniques used by the searchers to interact with the resource types. Examples of such actions, which are all identifiable in the notes, are "scanning a page", "entering a query", "adding a bookmark", and "selecting a link". There may turn out to be a finite number of possible actions that can be enacted on the World Wide Web at any time, but development of client software may come up with new possibilities.

"Need" signifies what currently is the searcher's information need, in the data this is mainly identified as a variable related to 'action', for example as shown above in Example 1: "reformulates query dahl+roald", or " Query 'usability'" in Example 3.

Two other properties that seem to be closely connected are *"results"* and *"relevance judgement"*. The former is used to categorise those observations made of what comes out of certain actions performed on web sites. Examples of "results" are "58 hits", "finds a conference report with overview of resources", "finds lots of translations", "no [hits on] R. Dahl", and "returns bibliographical result list". In Example 1 the line " long list that

contains entries on both roald and Dahl" exemplifies what we have decided to call a result, "Hits on only "jo"" from Example 2 is another result. In our third example we see results in the form of number of hits (58 hits).

Relevance judgements are observations that are made of the usability of the results, the utterance "too many", which refers to the 58 hits in example 3, being an example of such an evaluation. In the notes the following examples show the variety of observations related to relevance judgement: "not very interesting", "it's him - Nielsen", "this I don't understand", "ends up in a list every time", "strange that this should contain anything", "mmm – that was a nice one", and "service with restrictions". All of these observations are of utterances made by the searchers, but some notes are also based on the observer's interpretation of the searcher's interaction with the results. An example of the latter, which ends with an utterance from the searcher, is the following note: "[there were] many pages that were missing – [generating] frustration: "ugh, it looked so fine, but it was all just a mess"". The objects of relevance assessment varied. Sometimes it was a text, sometimes a list etc. It is probably necessary to differentiate between what objects this dimension should relate to.

The sixth property identified we have decided to term the *"information space"*. We use this to denote the expectations that the searcher has of the Web and how it should be navigated. The information space could be said to be the active part of the searcher's cognitive model of the Web. The information space seems to be partly built up on previous experience of the searcher and partly formed by what she expects to be found on the Web. The following two utterances exemplify an "inductive" and a "deductive" construction of the information space by the searcher respectively: "Dag og Tid [Norwegian newspaper] – we know that it contained an article" and "the DnB-group obviously doesn't have its own web pages". In the "Dag og Tid"-situation the searcher is triggered by the existence of that particular newspaper on the Web and quickly couples this with information from the printed version of the paper, which has contained an article on the particular topic. The "DnB"-situation reflects that the searcher originally believed that this organisation (a book club) had a web resource, a hypothesis that was "falsified" after a number of attempts at finding the resource, indicating a misconception in the searcher's information space. Other observations that indicate the existence of such a model include the following: "gets confused by [the design of] a commercial for Amazon.Com", "comments that there is so much [information] about R. Dahl", and "[I] know there are supposed to be book reviews here". Obviously this factor was also expected to exist in the data, reflecting, as it does, a cognitive model of the kind suggested by Ingwersen (1992).

The occurrences of comments that signify the use of specific *strategies* suggest that this may be a seventh property of Web interaction. The searchers make such comments prior to or during the interaction they "plan". Examples of strategy-signifying utterances are "I think I'll try somewhere else", "then I'll go to another author", and "[I'll] look for other newspapers".

It is unclear whether the latter category of observations should be treated as a dimension of its own, or whether we should see them as a reflection of the searcher's information space. We should also be aware of the existence of a conscious strategy, which we try to identify in pre-session interviews, applied by the searchers before actual Web interaction started. The strategy indicated in the notes can signify a change of the original strategy. For the time being we stick to the six dimensions, but hold our eyes open for more when we engage in a more systematic examination of our data. For now we look at them as six attributes of the Web search process. The second level of analysis was performed on the transcriptions of the video recordings.

4.4.6 Analysis of transcriptions of video recordings

The second stage of our analysis was performed on the recorded data. The three sessions lasted 1 hour and 44 minutes, 1 hour and 52 minutes, and 56 minutes respectively. With the help of a TV and a VCR and a lot of time and patience each session was transcribed into a table in Word. Each table row contains four cells where the following information was recorded: the *action* performed, the *time* spent in minutes and seconds, the *resource* visited, and the *utterances* made by the searchers. In the table we have also included a column tagged "WISS", the purpose of which we will explain in Subsection 4.4.7. The final column is also used for the observer to include comments about the searchers' behaviour and/or utterances (e.g. if the searcher mumbles, or has a humorous or sarcastic tone of voice etc.), and characteristics of the resource used that may be significant. From the analysis of the notes it became clear that this would be a suitable level of specificity for data recording.

The tables contained 456, 862, and 290 rows respectively, signifying the number of occurrences of individual actions performed on resource types. In Table 4.4 we show an excerpt from one of the transcripts.

In the notation we have used "." to denote a pause of up to 5 seconds, i.e. "....." signifies a period lasting between 25 and 30 seconds without the searcher uttering a sound. The

time has been recorded for each time the searcher entered a new resource. This seemed to be a point of specific interest as it was thought to signify a new turn in the session. We will later see how this observation was developed to identify a very important characteristic of our framework's core category.

In the description of the resource we have sometimes used short cuts to state the resource name, for example row 1 where the full name of the page, which is a sub-index in a subject index tree, is "Kvasir – kunst og kultur – litteratur – forlag". The context will make it possible to reconstruct the name of such abbreviated resource names. The searcher's comments (in italics) have been translated for convenience with the original text in parentheses, while other comments made in this column have been kept in English. Text in brackets is used to explain utterances.

The excerpt reproduced in Table 4.4 shows approximately two and a half minute of Web interaction. At this point of the search process the searcher has spent circa 50 minutes and investigated lots of different resources in order to find information about a predefined set of authors. The searcher is at this stage looking for information on books by the author Anne Karin Elstad and she has identified a few minutes earlier that Aschehoug is the publisher of this author.

In Table 4.4 we can see that 22 actions have been identified and that these include "scanning" or "looking at page", "selecting link", "backtracking", and "adding bookmark". In this respect the excerpt is quite representative, since these actions are very commonly used by all searchers. We also see that the searcher examines different parts of the resource type and she looks at the start page as well as a range of different indices. The only page that turns out to be of direct relevance for her is the short biography, which is bookmarked at the end of the shown sequence. Of the comments one should note the point she makes about it not being necessary to look at new books, as she believes that the author has not published anything recently. This could be said to represent the searcher's *work task knowledge*, which has found its form in her information space and which in turn prevents the searcher from performing "unnecessary" actions. Other comments represent the searcher's evaluation of the pages (results) that are presented to her.

Action	WISS	Time	Resource	Comments
Selects link	12	48.32	Kvasir - ... – litteratur – forlag	
Scans page	4		Aschehoug main page	. *hmm* .
Selects link	12		Aschehoug main page	
Scans page	4		Aschehoug – index to authors' biographies	. no entry on Elstad
Selects link	12		Aschehoug – index to authors' biographies	
Scans page	4		Aschehoug main page	..... *yes* (*ja*) .
Selects link	12		Aschehoug main page	
Scans page	4		Aschehoug – index to authors' pages	.
Selects back				
Looks at page, and points at link to "this autumn's books"	4		Aschehoug main page	*I don't see any point in entering* [the books of autumn] *really, I don't think she has* (*Jeg tror ikke det er noen vits i å gå inn der* [på høstens bøker] *egentlig, tror ikke hun*)
Selects link	12		Aschehoug main page	*Yeah, might check out* (*Ja kan jo sjekke*)
Scans page	4		Page presenting this autumn's books	*I don't think she has released anything this year so* (*Tror ikke hun har gitt ut noe i år, så*)
Selects link	12		Page presenting this autumn's books	*Could possibly try this one* (*Her kan en jo prøve*)
Scans page	4		Alphabetic index to authors of this year's books	. *no* (*nei*)
Selects back				
Selects back				
Scans page	4		Aschehoug main page	*It didn't work* (*Det gikk ikke*) .
Selects link	12		Aschehoug main page	
Scans page	4		Index to inexpensive books by Aschehoug	..
Selects link	12		Index to inexpensive books by Aschehoug	*Look at this* (*Se her da*) [the colour of the link indicates that she has visited this page recently]
Looks through top of page	5		Short biography and presentation of inexpensive books by Elstad	
Adds bookmark		51.00		

Table 4.4 Excerpt from session number 4

4.4.7 Results of transcriptions analysis

The findings from the analysis of observation notes were verified in the transcription analysis. We may call this the axial coding stage since the goal was to identify more categories and attributes and possible relationships.

The web search process can be seen as a series of "actions" performed on "resource types", where the other attributes in varying degree influence the choices made by the searcher. When dealing with the transcriptions we tried to find a suitable level of abstraction to classify the process. One such level was to transform the ISS model developed by Belkin and colleagues (Belkin, Marchetti & Cool, 1993) to suit web searching. Below we show how this could be done and discuss some of the strengths and limitations of such a model.

Web information search strategies (WISS)

Belkin's original model is presented in Subsection 3.1.3, and the development of the WISS model originally appeared in Pharo (1999). The idea behind the model was to see whether it was feasible to characterise web information search behaviour as a series of search strategies. We tried to apply the original ISS model to our empirical data, but it turned out that there were some limitations in this model with respect to the applicable dimensions. The values of some dimensions were also too limited. This made it necessary to develop the WISS model which is shown in Figure 4.3.

WISS	Method			Goal			Resource	
	scan	search	link	learn	select	illustrate	document	surrogate
1	X			X			X	
2	X			X				X
3	X				X		X	
4	X				X			X
5	X					X	X	
6	X					X		X
7		X		X			X	
8		X		X				X
9		X			X		X	
10		X			X			X
11		X				X	X	
12		X				X		X
13			X	X			X	
14			X	X				X
15			X		X		X	
16			X		X			X
17			X			X	X	
18			X			X		X

Figure 4.3 Web Information Search Strategies (WISS) model

We see that one of the original dimensions ("mode") has been removed and that some of the other dimensions are revised and/or expanded to make it fit for classifying web searching. It contains the following dimensions with their respective values in brackets: method (scanning, searching and linking); goal (learning, selecting, and illustrating and strengthening knowledge); and resource (document and surrogate). This model may be used to classify 18 different Web ISSes, which therefore have been called WISSes. Each row in our transcriptions was thought to represent the occurrence of a WISS and we classified them according to the scheme in Figure 4.3.

If we return to the WISS column in Table 4.3, we see that the excerpt consists of a searcher looking for information about books by a particular author. She alternates between scanning different indices in order to select a link that may take her closer to the target (WISS 4) and the selection of such links (WISS 12). In the latter case the goal dimension is classified as "illustrating" because the searcher already knows that the author has published a number of books and she wishes to find information about these books in order to represent (or illustrate) the author's production. This is also the reason why the last WISS is classified as 5 and not 1.

The greatest strength of the WISS model is its ability to perform a microanalysis of web search behaviour. We may identify each move, classify it, and then use statistical methods, such as log-linear analysis, or Markov models (Qui, 1993b), to look for patterns in each search process or in a large number of search processes. This would require a large set of data and a much larger amount of resources for data gathering and classification and analysis than was possible within this project's limitations. Another more fundamental problem with such an approach may be the meaninglessness of such a classification if it cannot be related to other explanatory categories. It may be that the WISS (and ISS) classification schemes are too abstract to say anything interesting about what actually happens during information searching. We see that characteristics concerning, for example, searchers' pre-knowledge, strategies, needs and information space disappear or drown in such a treatment of the data. For this reason we chose to look for more meaningful core concepts to explain what we believe are essential characteristics of Web search behaviour. The idea to look at Web interaction as a search process that alternates between source selection and inter-source navigation (search transitions) and interaction with "real" information sources (search situations) stems from the resource-dimension proposed in the ISS-model (Belkin, Marchetti & Cool, 1993).

The search process is made up by search situations and transitions

The excerpt represented in Table 4.4 describes a searcher interacting with the Web site of a Norwegian publisher. This can be seen as a quest for information about a topic, which in turn may be related to a search task. Prior to the interactions represented in Table 4.4 the searcher used two different sources which consisted of links to potentially relevant sources of the information sought by the searcher. Table 4.5 represents the interactions performed immediately before the data in Table 4.4. The last row in Table 4.5 and the first row in Table 4.4 are identical. Please note the same transcription techniques have been used as presented above.

We see that the searcher in the excerpt represented in Table 4.5 is trying to find an entry point for the publisher she is looking for. She navigates between two *meta-resources, or surrogates*, rather than in a "real" document. This signifies an important distinction in how people interact with the Web and possibly also with other information systems. Intuitively we got a feeling that this interplay constitutes an important part of the WIS-processes.

Inspects buttons, selects entry from bookmark	47.45		*Then I'll try to go to the publisher* (*Da prøver jeg å gå på forlaget*)
Scans page		Index on literature resources ("Litteraturpekere")	*Let's see then* (*Skal vi se da*)
Selects link		Index on literature resources ("Litteraturpekere")	
Scans page		Kvasir main page	
Selects link		Kvasir main page	**The previous link originally led directly to Kvasir's category on literature, but the link was apparently defect and the searcher is automatically led to Kvasir's main page**
Selects link		Kvasir – kunst og kultur	
Selects link		Kvasir – kunst og kultur – litteratur	
Scans page		Kvasir – kunst og kultur – litteratur	
Scans page		Kvasir – kunst og kultur – litteratur – forlag	
Selects link	48.32	Kvasir – kunst og kultur – litteratur – forlag	

Table 4.5 Excerpt (no. 2) from session 4

We choose to call interaction with real resources "search situations" and the interplay that goes on between two situations "search transitions". The dimensions that were identified above (in Subsection 4.4.5) and verified by the transcription analysis are in fact attributes of the search process' subcategories situations and transitions, except for the "relevance judgement" which logically must be a dimension of situations and not of transitions since the relevance judgements are made during interactions with potential information sources. In their turn search situations and search transitions are dimensions of the *search process*, which is the core category in our data.

In our third level of analysis we have tried to see how our core category relates to the other categories ("selective coding"). In addition we continue looking for new dimensions of our categories and trying to identify inter-category relationships. For this analysis we have used stories that are based on the transcriptions and real-time notes, as well as the questionnaires and the pre- and post-search interviews.

4.4.8 Story analysis

Based on the studies of information seeking and retrieval literature and our own experience in using the Web, we anticipated particular factors to play a role during Web searching. As stated in Chapter 2 we have a task-based focus on web searching, thus *work task* as well as *search task* were thought to play a role in Web interaction. This necessarily led to a biased view of web interaction, but as we have stated previously we are interested in problem solving on the Web, which means that the searcher must necessarily have a problem or task to begin with.

In addition, characteristics of the *searcher* and her *environment* are thought to be of importance during the process and were presupposed to be identifiable in the data.

The questionnaires were helpful in stating the initial problem that led to the searchers' engagement in Web searching. In addition, they provided us with information about the individual searchers and the context within which they worked.

We have used all different kinds of search session-related data to make stories tell what happened in the individual sessions. The stories focus on the transitions and situations and their attributes and the relationships between these categories[17] and their attributes as well as relationships with other external categories and their attributes. The stories were

built up in a way that set the context of each session in the beginning and the session was then described in natural terms. If necessary, terms that represented the category that appeared were added to explicate the occurrence. A sample from one of the stories is shown below. Text in Italics represents interpretative parts of the sample. We have coded the attributes with names in Italics either embedded in the text or in parentheses and used arrows to indicate the identified relationships. We have also tried to identify whether attributes affect one another positively or negatively, the occurrences being coded with plusses and minuses respectively. We will explain how to classify negative and positive relationships later. In the excerpt below we can see that accumulated results positively affect the information space twice and that the information space twice negatively affects actions. In the "accumulated results +→ information space" cases the relationships are used to indicate that we can identify an occurrence of the searcher's model of the Web being strengthened by the accumulated results. In other words it signifies that a resource's value as experienced by the searcher at a previous stage makes her employ the same resource again. In the cases where information space negatively affects action we have used the relationship to indicate that the searcher's model of the Web did not foster particular actions that could have led her to useful parts of the resource.

> The searcher first checks out a bookmark (*action*) to see what page it leads to. It's apparently a page in German on Norwegian female authors (*resource type*), which she found in the previous session (*accumulated results*). This may be seen as a situation, but it is not a part of the *ST strategy* she has decided to follow. It probably is a distraction, because the next action (starting transition 1, or session 4 transition 1 - S4_T1 – 44 secs.) she takes is to select another bookmark (*action*) leading her to an index to different literature resources (*resource type*), which is a page that all the students working on this task together have decided is a valuable resource (*actors → accumulated results +→ information space*). She traverses and selects the Norwegian subject index Kvasir's category on literature-publishers (*action – resource type*). In Kvasir she selects the link (*action – resource type*) to the publisher Aschehoug which initialises the first situation in this session (situation 1 or S4_S1 (session 4 situation 1).
>
> *Throughout the session the index to literature resources is used as a landmark, the use of which often signals the start of a new situation (information space). In S4_T1 it is used to lead the searcher to the Norwegian subject index Kvasir, which in turn is used to find the publisher Aschehoug. The link initially followed from the index to literature resources originally led directly to Kvasir's literature index, but this had not been updated when the Kvasir hierarchy was restructured - thus the searcher had to get to this page manually (technical problems - → information space). Thus S4_T1 could have been performed even more smoothly if the landmark page had not pointed to a non-existent page.*
>
> The publisher Aschehoug's web resource (*resource type*) is the source used in situation S4_S1 and it is obvious that the searcher is aware of its existence (*information space),* the colour of the link also indicates that she has visited it recently (← + *accumulated results*). The searcher

[17] Although "search process" is the super-category, we focus on its core dimensions, situation and transition.

states that she is looking for information about the author Karsten Alnæs (*ST goal*). Within the publisher's site she first selects a link (*action*) to a short index on authors with their own web pages (*resource type*). Since Alnæs is not among them (*relevance judgement*) she quickly identifies and selects (*action)* the page listing new fiction where she looks for books (*action – resource type*) by author Alnæs. There is however no entry on him in this page. The searcher therefore scans through a page listing all new books, but when this also is fruitless she gives up (*relevance judgement*).

No querying is involved in this situation. The searcher finds a well-known source (she also used it in the previous session) through traversing links in a subject index (action – resource type). Her navigation within the source seemed well prepared and targeted and is obviously in accordance with her ST strategy. She quickly gave up when she realised that this author was not published lately by the publisher (accumulated results → ST strategy). This indicates that the searcher didn't know who published the author (task knowledge) and suggests perhaps that she had not done any background research on this. It is also interesting to notice that the searcher only looks for information on books published recently and is seemingly not interested in looking for information about the author from other angles (which indicates a deficiency in her information space - → action). Looking in the publisher's list of authors could, for example, be successful. The situation lasts 1 minute and 33 seconds.

The searcher looking at her bookmarks and then backtracking (*action – resource type*) to the Kvasir page that links to Norwegian publishers signifies the transition (S4_T2) to a new situation. (16 secs.) The searcher selects a link to Gyldendal and starts the next situation.

The preceding transitions leading to the first two situations differ in the sense that in S4_T1 the searcher first needed to find the Kvasir page containing the links to Norwegian publishers. The searcher's accumulated efforts → therefore save the searcher time and effort in the remaining transitions. On the other hand she could have bookmarked this page and saved herself even more time because then she wouldn't have to backtrack every time she needed to use it.

Situation S4_S2 is very similar to its predecessor (*ST strategy*). The searcher explores the publisher Gyldendal's web resource (*resource type*) where (based on her *information space*'s representation of publishers' resources) she selects the link (*action – resource type*) to a page presenting novels and short stories published "this year", but without finding the author Alnæs (*relevance judgement*). She thereafter scans through the main page of the site (*action – resource type*) and finds and follows a link to another publisher (S4_T3 – 3 seconds) "Tiden". Situation S4_S2 lasts 1 minute and 5 seconds.

If the searcher had explored the site more thoroughly (information space - → actions) she would have found the publisher's main catalogue where it would be reasonable to look for entries on Alnæs.

The start page for "Tiden" (S4_S3) (*resource type*) only contains information about the publisher's address and a link to a page on fantasy literature. Having scanned (*action*) the page for approximately 20 seconds to judge it irrelevant (*relevance judgement*) the searcher therefore chooses to backtrack (*action*) to Gyldendal's main page (S4_T4 – 3 seconds). It is clear that the page does not meet the searcher's expectations concerning its content and structure (i.e. her *information space → resource type*).

This situation was apparently not generated with the basis in the original ST strategy. On the other hand it exemplifies that ST strategies are dynamic: "if the original site does not contain pages that directly fulfil my need, then scan it for links to external pages that might".

The next situation - S4_S4 - is performed within Gyldendal's web pages (*resource type*), and the searcher now looks for information on the author Roald Dahl (*ST goal – ST complexity*). This is somewhat surprising, as the searcher claimed that she would deal with each author separately and she has yet to find anything on Alnæs. It probably indicates frustration over not finding the expected information on Alnæs (searcher feels that it has been a waste of *time* and *effort* - → *motivation* → change in *ST strategy*). This time she chooses to start her quest in the publisher's main catalogue of books (*resource type*), not specifically on new releases, raising the question of why she did not follow this strategy in the preceding 2 situations (*attention*). In this page she finds a query form where she enters "dahl, roald" (*education* +- → *search knowledge* +- → *action – resource type*) [excluding the quotation marks], which returns a results set that she systematically explores (*action – resource type*) - by visiting the pages in the hit list in descending order. She finds two pages presenting individual books by Dahl and one that presents several books by the author, these are all bookmarked (*relevance judgement*). When she has scanned through only page one (*action – resource type*) of the three page result list - i.e. 10 out of 30 hits - she decides to stop searching for the author in this source, saying "I don't think there's more [here]". The entire situation lasts 3 minutes and 6 seconds. She selects a bookmark (*action*) to end the situation.

The searcher also in this situation takes advantage of her accumulated efforts by re-using the same resource in looking for another author. In the following situations the searcher continues to look for information on the same author. This might indicate that her success in finding relevant pages serves as an inspiration (accumulated results + → motivation) in continuing to look for information on the same topic rather than returning to the previous, seemingly harder-to-find author. It should be noted as an interesting factor that the searcher very quickly decides whether a page is relevant or not, having presumably to do with the WT goal which affects the relevance level and thus the relevance judgement. It is also interesting to note that to deem a page non-relevant is often done at the query results level, without actually looking at the page. As we shall see this is repeated throughout the entire session. The fact that she only looks at 1/3rd of the retrieved pages is interesting. This might be explained by the occurrence of a duplicate in the result list. She decides to stop when the first retrieved page occurs a second time among the 10 first pages, which may have caused the searcher to reason that the remaining links will point to duplicates as well (search knowledge + → action).

After conferring with her notes the searcher now decides to start looking for book reviews (*ST strategy*). For this she uses another "landmark" as her point of departure, which is an index to Norwegian papers' literature reviews. As with the previously used landmarks this has also been established "in community" (*actors → accumulated results ← information space*). In transition S4_T5 she scans this book review index (*action – resource type*) and selects the link (*action*) to Aftenposten. (20 secs.)

We end our sample here. The session continues with a total number of 76 transitions and 64 situations.

The sample contains examples of a variety of attributes that influence the search situations and transitions. We can also find some examples of attributes that do not explicitly relate to situations and transitions, but that are characteristic of the searcher and her environment. We see the occurrence of *actors* within the searcher's *social environment* and we see that the *searcher* is characterised by her *motivation*. A third non- situation or transition related attribute is the *strategy*, which can be a more or less conscious and explicated characteristic of the *search task*.

During this coding stage we understood that it was necessary to refine the relevance attribute. The relevance can be judged according to the needs of the searcher, but there are also situations where a more objective relevance *level* is used for judging. We have seen that searchers collect potential information on the Web aimed at satisfying the needs of a third party, for example pupils who work on a project on Norwegian authors. In such a case the target group is not the searchers themselves. This situation can be compared to the work of reference librarians who collect information for a patron. Saracevic's (1996b) distinctions between different levels of relevance seem to be fitting for the measurement of *relevance level*. If the searcher's role is that of an intermediary who is not in direct contact with her client during the search, which relates to the example above, she will measure the relevance of the retrieved documents at the topical level. On the other hand, when the searcher looks for documents that may help her solve a problem, or execute a work (sub)task related to her own needs the relevance level is situational.

Two other characteristics, which seem to play significant roles particularly towards the end of search sessions, are *time* and *effort*. We shall return to discuss them later.

The searcher's knowledge about the work task that she performs we have chosen to call *work task knowledge*. *Search knowledge* has similarly been attached to describe her knowledge in information searching.

Work tasks may affect WIS-processes in many different ways and by so many of their features that it would be impossible to list the possible attributes a priori. During the process analysis, however, we may identify some of the features that were expected. The goal/purpose of the WT would be one such attribute, while others that were identified include WT size and complexity. The latter two attributes' roles are, however, much more indirect, and do, in particular, affect the search task. Similar characteristics of the work tasks generating the sessions that we have studied may well have led to this conclusion, since they are all of similar complexity and size. Nevertheless, it seems that these features influence the searcher foremost through her choice of search task strategy.

The searcher generates *search tasks* in order to support her handling of the work task. The attributes of search tasks are to a certain degree similar to those "assigned" to the work task. We shall examine them more thoroughly in Subsection 4.5.4.

We believed that search knowledge was an important attribute and discovered that it actually was necessary to divide it into a general attribute and a search *system* knowledge attribute, where the latter represents the knowledge about the client program's features. In addition "search task strategies" is used to explain how the searcher has decided to attack the search task. ST strategy has not been further sub-divided, although that clearly may be done given suitable data.

The searcher interacts with a *social* and/or *organisational environment* which consists of other *actors*, but which can also be characterised by its *domain* and more or less formulated *strategies* and *goals*. We present below a sample that exemplifies the *goal*, *size*, and *complexity* of a work task as well as the *strategy* and *goal* of a search task. The sample is taken from a session where two searchers perform a session together.

> The searchers are working on their final year thesis at the education in library and information science (LIS) at Oslo University College (*domain*). The topic of their thesis is the Norwegian author Jo Tenfjord about whom they were asked to write a bibliography that also contained a short biography (*WT goal*). The thesis represent ½ half semester (5 credits) of work (size) and is required to achieve a certain academic level both in content and format (*WT complexity*). At the time of searching the task is quite new and the searchers feel stressed for a variety of reasons, among those being the fact that the author herself is still alive and the thesis might be published for an anniversary celebrating her 80th birthday (*motivation*).
>
> Prior to this session the searchers have searched the Norwegian national bibliography for literature by and about the author and they have also sporadically used the Norwegian subject index Kvasir to look for information on the author (*accumulated results + accumulated efforts*). Attempts in the latter case have not returned any relevant results.
>
> The searchers both regard their Internet and general computer experience as medium. The searchers both use email and Web and one of them also claims to have experience in using FTP. They use the Web to "search for information". One of the searchers has home access to Internet (*search knowledge*).
>
> Both searchers believe that the Web will be a source of information of medium importance in the work with this particular task. More important sources are "lecturers at the department" (*actors* in *social environment*), bibliographic databases, books, journals, and newspapers.

> One of the searchers has information science as her special subject. This means among other things that she has more training in searching databases and formulating queries than students without this specialisation (*search knowledge*). The other searcher's speciality is in literature sociology. She has had 120 hours training in information searching which is standard for all students at the LIS faculty (*search knowledge*) at Oslo University College. Through her specialisation she has received extra *task knowledge* about Scandinavian literature and interpretation and mediation of texts. This may of course be of help when looking for information about a Norwegian author.
>
> The searchers have a very open-minded point of departure for the session, intending to see if there is anything about the novelist on the Web (*ST goal*). They are aware of some organisations that she has been involved with and would like to see if there is any mention of her there. In addition they have planned to search the Danish and Swedish national bibliographies via the Web (*ST strategy*).

In Subsection 4.5 we shall systematically examine the categories that we have identified and organise them in the format of a conceptual framework. The result of this constitutes our method schema's domain.

4.4.9 Summary of the pilot study's methodological parts

The pilot study can be divided into different stages in the sense that a variety of different kinds of empirical data were analysed. We shall summarise the pilot study's methodological parts by focusing on how we used the different data.

The first data were those stemming from the interviews with the searchers which provided us with background knowledge about the topics and goals of their theses. This knowledge has been used to supplement our analyses at all levels, but played a minor role compared to the data generated by the search sessions that we observed and recorded.

The second stage in the study consisted of the analysis of observation notes made at the time of searching. This was a very influential stage, in the sense that it was here that important characteristics of the sessions attracted our attention. The notes generated a set of attributes that represent different dimensions of the WIS process.

In the third stage we transcribed the recordings of search sessions, which led to the identification of the Web Information Search Strategy (WISS) model influenced by Belkin and others (Belkin, Marchetti & Cool, 1993; Belkin et al. 1995). We have, for the time being, put the WISS model aside due to the costliness of data collection and analysis and because we believe the WISS model needs to be combined with

other explanatory categories to produce meaningful data. We therefore chose to go further with one of the dimensions of the WISS model, the resource type, which gave us the idea to consider WIS processes as an alternation between interaction with meta-information (surrogates) and interaction with documents that the searcher thinks may contain information. This idea led to the identification of search situations and transitions.

In the fourth stage we created stories representing the WIS processes. The stories were analysed using the categories and attributes that had been identified during the analysis of previous models (presented in Chapter 3) or in the previous stages of the pilot study. Additional attributes turned up during the story creation, which were also used in the analysis. The fifth and final stage of the pilot study would be the creation of the domain description, which we present in the forthcoming subsections of the present chapter.

We found this methodological approach very fruitful because it made it possible for us to explore the effects of the different categories and attributes on the WIS process at the same time as we aimed at understanding the nature of the categories from a WIS process point of view. This was particularly important when examining the categories that are not well defined in the LIS literature, which is the case for both "work task" and "search task". We kept the category definitions open in order to see in what categories the data naturally fit. This exploratory approach improved our understanding of the categories and how they affected the WIS process.

4.5 The domain description

We have identified five broad categories or concepts that seem to be sufficient to describe what happens during task-based WIS-processes. The five categories are the *work task*, the *searcher*, the *social/organisational environment*, the *search task*, and the *search process*. The search process in turn consists of two subcategories; search situation and search transition. We believe these categories constitute a conceptual framework that can be used to explicate the domain of our research problem.

4.5.1 Work task

As explicated by Hansen (1999) and Vakkari (2003) there has been a lack of consistency in the literature concerning the use of the term "task", which can be used to refer to "work task" as well as "search task". We prefer to use these terms to avoid confusion. A

work task is a series of actions in pursuit of a certain goal, where the performance is seen as an integrated part of the task.

The work task is strongly related to "problem" or "problematic situation" (Wersig, 1971), but we believe this term to be more inclusive. Although problematic situations clearly activate information needs which in turn may generate search processes there are also less "cognitive" and individual-centred situations that may lead to search processes. We have decided to use the term work task in accordance with Byström (1999).

The work task can be treated from the viewpoint of the searcher, i.e. the perceived work task or it can be viewed as an objective entity, i.e. the task as assigned to the searcher. We have the perceived work task perspective, focusing on the searcher's interpretation of her job, which is compatible with observing searchers in WIS processes where they interpret their WT goals and perform in order to execute them. See Byström (1999) for a discussion on the relationship between subjective and objective work tasks.

Important attributes of a work task include goal, complexity, size, and stage. We shall discuss these attributes below.

The purpose or *goal* of the work task is what is intended to be its final result as perceived by the task executors. The goal may be vague at the start of the work task process and be sharpened as the process progresses. Some tasks have a final product as their goal, for example when an article is accepted by a journal some particular task may have reached its end. Simpler tasks, such as collecting a cup of coffee, are, if the facilities provide for it, easy to solve. Other tasks do not have a proper end and may be the reason for creating organisations to deal with them. Very simply we can say that one of the tasks of the Faculty of Journalism, Library and Information Science at Oslo University College is to educate librarians for Norwegian libraries. This exemplifies a task that does not end in the foreseeable future since there is continuous demand for this kind of labour in Norway.

Järvelin (1986) refers to different types of tasks decision-makers engage in which demand different kinds of results. He claims that different task goals require different demands on information processing (Rich, 1983), type of information, information sources, and information search strategies. We believe that this attribute is an influential factor during WIS-processes, perhaps more indirectly in its influence on search task goals.

The work task *complexity* is used to describe the combination of the

- dimensionality of the task

- predictability of task performance
- number of sub-tasks involved

Task complexity is individually perceived and depends on the searcher's task knowledge. Byström (1999) has investigated how task complexity affected searchers' needs and choice of different information types and sources. From her findings it seems plausible to hypothesise that only tasks of a certain complexity trigger the use of databases and other complex information search systems.

The *size* of the work task can be measured in terms of the number of working hours assigned to it.

Work task *stage* is related to the stage in the information search process proposed by Kuhlthau (1993) and Ellis' categories, which were both discussed in Chapter 3, as well as the problem treatment process of Wersig and Windel (1985), who have described how different factors influence people's behaviour towards problem solving. They describe the "problem treatment process" (p. 15) starting with a human being at a certain time, being influenced by the factors: present situation, past states of the organism (personal history), system of preferences (values, opinions, attitudes, etc.), and set of potentials (cognitive, affective, aesthetic, etc.). The problem treatment process consists of up to 12 states, but Wersig and Windel are very clear in claiming that external (information) searching behaviour does not necessarily take place at all these stages. In fact they state very clearly that information "may play a role but not necessarily a central one" (Wersig & Windel, 1985, p. 15).

In addition to the attributes mentioned, the work tasks certainly have other attributes which we have not examined here. We have, for example, not explored factors relating to the numbers of actors involved in a task or the economic as well as other resources dedicated to the task.

4.5.2 The searcher

The searcher is the person who interacts with the systems. More than one searcher may be involved in a search process, although usually only one of them performs the actual actions. In the latter situation the decisions about which actions to perform will be the subject of discussion among the searchers.

We believe that the following eight attributes of the searcher are of special significance: work task knowledge, search knowledge, search system knowledge, education, motivation, tenacity, uncertainty, and attention. These are discussed in more detail below.

The *work task knowledge* is used to denote the searcher's pre-knowledge and prejudices concerning the task(s) she tries to perform by interacting with the system. Such knowledge is the result of the searcher's previous performance of similar tasks, her work with current tasks, as well as knowledge acquired from reading about relevant topics and discussing it with others. Vakkari (2001) argues that work task knowledge affects the searcher's ability to make relevance judgements and formulate queries.

The term "theoretical sensitivity" (Glaser & Strauss, 1967), which refers to a researcher's personal knowledge of a field based on his professional and personal experience seems usable to characterise the context in which task knowledge is built. "Task knowledge" should be used specifically for knowledge related to the task that the searcher is involved with, but the borders between task knowledge and theoretical sensitivity may be unclear especially if the former is vague. Several authors have investigated how conceptual, domain or subject knowledge influence user-intermediary interaction (Ingwersen, 1982), information searching in on-line catalogues (Hsieh-Yee, 1993), hypertext systems (Marchionini, Lin & Dwiggins, 1990), and the Web (Hölscher & Strube, 2000), and it is clear that such knowledge strongly relates to task knowledge. When focusing on the task knowledge we wish to draw attention to that particular part of the domain knowledge and this focus may also help to reveal how a lack of task knowledge can influence searching.

Search knowledge covers the searcher's knowledge in searching information systems, including the use of query language operators, being aware of synonyms, knowing about the implications of searching in specific fields versus free-text searching etc. There exist a variety of studies of the way search expertise influences search execution in different kinds of information systems (e.g. Ingwersen, 1982; Marchionini, Lin & Dwiggins, 1990; Marchionini et al. 1993). Hölscher and Strube (2000) focus on how search and background knowledge affect Web searching. Search knowledge may be of a theoretical origin as well as being based on previous information searching experience. Search knowledge gained from using one system may be valuable both as inspiration in how to formulate a search in a new system, as well as in building up the searcher's self-reliance with respect to how to "attack" the system.

A specific kind of search knowledge is the searcher's *search system knowledge*. This knowledge is specific for the system in use and more specifically has to do with the

searcher's knowledge in using the Web client software. The software may provide different features to enhance the efficiency of Web interaction given that the searcher is aware of them. The searcher may take advantage of a pre-defined list to resources provided by the software, she may use its bookmark list, she may use the reload option to get the latest version of individual pages etc. This attribute emerged from the pilot study.

The searcher's *education* has been treated as a factor in many studies (e.g. Ellis, 1989; Ellis, Cox & Hall, 1993). It should include the influence of all levels of education received and courses attended by the searcher. Education probably influences the work task knowledge, although this relationship may be implicit. The searcher's education is probably also related to the domain of her social and organisational environment in the cases where the work task is related to her job.

Motivation refers to the searcher being influenced to engage in interaction generated by the results presented to the searcher by the search system (Web). Both motivation and tenacity (below) are attributes that were identified as a result of the pilot study.

Tenacity is used to describe the searcher's inclination to invest time and effort in "impossible" or very repetitive and boring tasks. This attribute is particularly important when an episode of interaction takes a long time and it "prevents" the searcher from interrupting the interaction.

Kuhlthau (1991; 1993) focuses on how *uncertainty* is related to different stages in the search process, but it can also be related to instances of uncertainty that appear during information searching. We use it in this latter sense, since we want to focus on the micro-level to see how instances of uncertainty influence the search process.

The searcher's *attention* is used to describe her ability to notice useful information as it appears during the search process. This attribute evolved from the pilot study.

Other factors can be used to characterise searchers, for example background, age, gender, cognitive style, experience etc. We have made our selection of attributes either because they play influential roles in the data we analysed in our pilot study or because they have been judged important in the literature we have examined.

4.5.3 Social and organisational environment

This category is emphasised by many IR&S model developers as an important part of the searcher's surroundings, in the sense that the searcher constructs her understanding of a (work) task via interaction with her surroundings. Discourse analysts (Frohmann, 1994) focus on how specific institutions give meaning to the core elements discussed within a field or an organisation. It is also relevant to point out the social functions of paradigms (Kuhn, 1996), and how normal science is performed within a set of boundaries defining what kinds of problems are to be studied and what methods are to be used.

Previous research (Rasmussen, 1990; Ingwersen, 1992) has pointed out the influence of the environment's *domain* on information behaviour, Hjørland and Albrechtsen (1995) even propose that it should be the main unit of study in information science. The domain will change as the setting of the work task changes; for example different domains (and environments) will be involved when the work task relates to the searcher's job in a computer firm as opposed to when it has to do with her involvement as leader of the community's chess club.

The *actors* refer to the persons involved in the searcher's social/organisational environment. This may include colleagues at work, members of an organisation, friends, or others that in some way may be involved in surroundings that have an interest in what the searcher does.

The goals and *strategies* of the environment/organisation refer to its reasons for existence. Organisations have their own etiquette as to how tasks should be solved. Written and tacit procedures exist which the searcher must take into account when dealing with problems/tasks of different kinds. Audunson claims that such "norms, rules, and structures" explain why "decision makers often gather information without using it" (Audunson, 1999, p. 526).

Ginman (1987) has examined how personal and corporate dimensions affect chief executive officers' information behaviour. She suggests a number of attributes that could give a richer perspective to the social/organisational environment. Such possible attributes include organisation age, organisation size, and organisation life cycle.

4.5.4 Search task

The search task can be characterised by its goal, its complexity, and the search strategies used to execute it. Reid (1999) suggests that "IR tasks" should be the focus of a new

paradigm for information retrieval. A search task, in some instances, can be identical to the work task, for example for an intermediary whose work task is to perform a search for a client and submit the result in form of a printout, a list of references, etc.

The search task *goal* is what is intended to be its final result i.e. what information, documents, facts, etc. the searcher understands to be the desired result. The ST goal may be vague at the start of a search process and may be sharpened through interaction with the system. In addition to the ST goal's topical content, search task goals can, according to Marchionini (1995), be characterised by their specificity, quantity, i.e. the amount of information needed, and timeliness. The latter factor also relates to how urgent the information is needed.

With Järvelin we can say that the search task goal may be to collect problem (task) information, domain information, or problem (task)-solving information (Järvelin, 1986; Byström & Järvelin, 1995).

The *complexity* of the search task refers to its multidimensionality; i.e. the variety of subtasks within the search process. The subtasks may reflect a number of similar search subtasks, for example when a searcher looks for information about books by five different novelists during a search process. It may also have something to do with the variety of angles used by the searcher to attack the search task, for example when a searcher looks for biographic information about a novelist, including as many dimensions of her life as possible.

One way of examining search task complexity could be to study the consistency in selection of search concepts and search terms across time (Iivonen, 1995) as well as how many different search concepts that are used for the same search task.

Byström's (1999) analysis of *work* task complexity might also be useful applied to *search* task. Thus search task might be classified by analysing the determinability of the information needed, the process, and the outcome of the process.

Search task *strategies* constitute the searcher's plan for executing the search task. ST strategies may include specification of which sources to use and the techniques that should be used in order to explore them. The strategies may also state in what order sources and "queries" should be explored and executed. Examples of initial strategies could be:

"Depth first" in a predefined set of sites

The user searches for information about one topic at a time in several different web resources before changing to another topic. The resources are selected from a predefined set i.e. the user is in some way familiar with the resources. She may have used them previously or her attention has been turned towards them for some other reason (good reputation, famous brand name etc.). When a user uses a search engine, the hits may of course consist of links to unknown resources. The point here is that the point of departure is a previously known site.

Width first in predefined set of sites

The user searches for information about several topics in one web resource at a time before changing to another resource. As above, the resources are selected from a predefined set.

Web-berry picking

A special kind of explorative searching we may call "web-berry picking" – after Bates (1989). This is used to describe a strategy where the searcher traces links from page to page with the idea that if the present page is judged to be relevant then the pages it links to also might be relevant. The searcher 's knowledge will change as she interprets the web page, in turn influencing which link she decides to follow.

Exploring unknown territory

A user may also try to find information by using resources that she is not familiar with or by using known resources in an unfamiliar way. This may be characterised by much trial and error, but it may be very valuable as a way to get acquainted within a field and to build up a collection of resources.

4.5.5 The search process

As previously stated we see the search process as consisting of search transitions and search situations. A search process is a series of transitions and situations and switches between them, following the three basic rules:

1. a search process always starts with a transition
2. a transition may be followed by a transition or a situation
3. a situation is followed by a transition or by the end of the process

The end of the process may be controlled or uncontrolled. In the first case the searcher exits the client program using any of its exit-commands, while in the latter case the program crashes and the searcher does not choose to resume the process. If the searcher restarts the program and continues working with the search task, the process should be treated as a whole, not two separate processes. The definitions of the attributes are, to a large extent, derived from the data analysis.

Search transitions

Search transitions are executed in order to find resources in which the searcher believes there may be information that can help her execute her task. We have stated above that transitions consist of source selection and inter-source navigation, but a third way of explaining it is to say that the transitions deal with *meta*-information.

Search transitions share many characteristics with search situations, which we examine below. The main difference is that the transitions take place on the meta-level – i.e. the searcher interacts with information surrogates with the intention of finding resources that she believes may help her solve her task. Thus a transition can be compared to an information-seeking strategy (ISS) (Belkin, Marchetti & Cool, 1993) performed in a meta-information resource. As the searcher has no contact with potentially task-solving resources during transitions, no direct relevance judgements can be made. Of course the searcher will decide which links to follow or reject in a query result list or a subject index, but as soon as the link is selected the situation starts. Thus all relevance judgements during transitions are based on surrogates and lots of rejections take place "silently".

Action is used to describe the moves (Fidel, 1985) made by the searcher during a transition. These include "following link", "entering queries", "reading a page" and possibly others – the definition is based on the findings in our pilot study. There are additional client-program-dependent actions, but we have not found any other types of actions independent of the software used to interact with the Web. It is possible to identify certain entering or leaving *actions* when analysing search transitions – such actions of course also signify the start/end of the next transition or situation.

The *accumulated results* refer to the information already found. This includes information found in previous transitions and situations as well as information found in the current transition. We can say that the search process is characterised by the searcher constantly stacking (usable and useless) results into a pile and choosing to use (or not to use)

whatever is in that pile in the current transition. It may also refer to results accumulated during preceding search sessions and through other work sub tasks.

The *accumulated efforts* refer to how much work the searcher totally has had to invest from the start of the session to the current position. In addition it can refer specifically to effort invested in the current transition.

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 her knowledge about the Web and the existing resources on the Web, but also on her knowledge about institutions and organisations that she expects to be represented on the Web.

Tauscher and Greenberg (1997) support the notion of such a model. They have examined revisiting-patterns in Web resources and found that 58 per cent of the pages a user accesses are revisits. Thus the searcher has built up a reservoir of resources that forms her map, or information space, of the Web.

Time can be used to specify how the total amount of time spent during a search process influences the current transitions, but it can also relate to the specific time used in that transition. Time is clearly related to effort and as time goes so will the accumulated effort logically increase. Time is also an important factor when it comes to choosing what resource is used in order to solve a task (e.g. Byström, 1999). Similarly on the micro-level how much time is left may influence the searcher's choice of what Web resource to use.

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

There is a myriad of *resource types* available on the Web, which differ with respect to content and format. In addition to document types known from the world of paper-based publishing, such as newspapers, scientific journals, dissertations, novels, and collections of poems there are many new genres that have originated on the Web. All kinds of institutions and organisations have developed resources to present themselves, combining characteristics of the organisations as they are presented in the real world and features that are specific for the digital format. Examples of the latter could be the use of chat rooms and various forms of discussion forums and interactive features. Clearly these resources can be used during both transitions and situations, depending on the search

task. In most transitions, however, we believe that resource types like "search engines", "subject indices", and various forms of "linked lists" play the major role.

The searcher may well change between different resource types in a single transition. She may for example inspect a search engine's start page and from there go to its help page, which are clearly two different resource types although published by the same site. Similarly the search engine's query form requires different interaction techniques and constitutes a different resource type compared to the result of a query using this query form.

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

The search situation

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 to her to execute her 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.

Situations and transitions share many attributes and below we re-use several of the definitions from above. A situation can be compared to an ISS (Belkin, Marchetti & Cool, 1993) performed in an "information resource".

Action is used to describe the moves (Fidel, 1985) executed by the searcher during a situation. These include "following link", "entering queries", "reading a page" and possibly others. Please note that "entering queries" here would refer to such actions performed within a resource's local database or search engine, for example a newspaper's database on its articles or the use of the client program's "find"-option, usually found in the "edit"-menu. The searcher will often switch between actions during a situation, depending upon what actions are possible and necessary to interact with the resource types in use. The discussion under transition-actions is also valid for actions performed during situations.

The *accumulated results* refer to the information already found. This includes information found in previous transitions and situations as well as information found in the current situation. We can say that the search process is characterised by the searcher constantly stacking (usable and useless) results into a pile and choosing to use (or not use) whatever is in that pile in the current situation.

The *accumulated efforts* refer to how much work the searcher totally has had to invest from the start of the session to the current position. In addition it can refer specifically to effort invested in the current situation. The effort invested will logically increase over time, thus a clear relationship exists between this attribute and the time-attribute.

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 her knowledge about the Web and the existing resources on the Web, but also on her knowledge about institutions and organisations that she 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 invested in a situation. Time, as stated above, influences effort. Time is also an important factor when it comes to choosing what resource is used in order to solve a task (e.g. Byström 1999). Similarly on the micro-level how much time is left may influence the searcher's choice of what Web resource to use.

Relevance judgement relates to the searcher's evaluation of the pages found, which may be of use to her in different degrees. We do not state any predefined categories for relevance judgements, whereas in other studies binary (relevant or not relevant) or ternary (adding "partially relevant" to the former two) relevance measures have been used. See Borlund (2000, p. 35-37) for an examination of degrees of relevance.

By *relevance level* we mean that the criteria used for evaluation may be related to the work task, which is what Saracevic (1996b) calls situational relevance, but they can also be related to other levels, e.g. when an intermediary judges the relevance for a (potential) user. The relevance judgements are also taken in accordance with the organisational preferences, thus socio-cognitive relevance (Cosijn & Ingwersen, 2000) may also affect the judgements.

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

There are many kinds of *resource types* available on the Web, which differ with respect to content and format. The searcher may well change between different resource types in a single situation. She may for example inspect a publisher's author index and from there go to a page presenting a specific author, clearly two different resource types although published by the same site. Please refer to the discussion on resource types made under the transition heading.

The term "*Technical problems*" is used to describe problems caused by the software in use, which are the same problems that we discussed above.

We do not state that these attributes are the only ones characterising the search process, but during our pilot study we found these to be the ones standing out in the data. As discussed in Subsection 4.1 on method schemas we have not aimed at identifying any category or attribute involved in, or affecting, WIS processes. Our intention is to give the broad meaning of important categories, while leaving open enough to make it possible to recognise in the data traces of phenomena that represent them.

4.5.6 Relationships

Figure 4.4 (next page) is a representation of the framework's categories and the relationships that may exist between them.

The attributes that we have discussed above are listed within the boxes. In addition each box could have been equipped with a general "other" attribute. Here we shall only summarise the relationships on a general level based on what we have found in the literature and in our pilot study. For many of the relationships the summaries will be brief, since further research is necessary to learn more about them. We have not explicated the "inter-attribute" relationships in the figure since this would give too complex a picture and because we cannot state all possible relationships. For a more thorough examination of such relationships we refer the reader to the case presented in Chapter 5. The numbers are used to identify the different relationships and are referred to in the text.

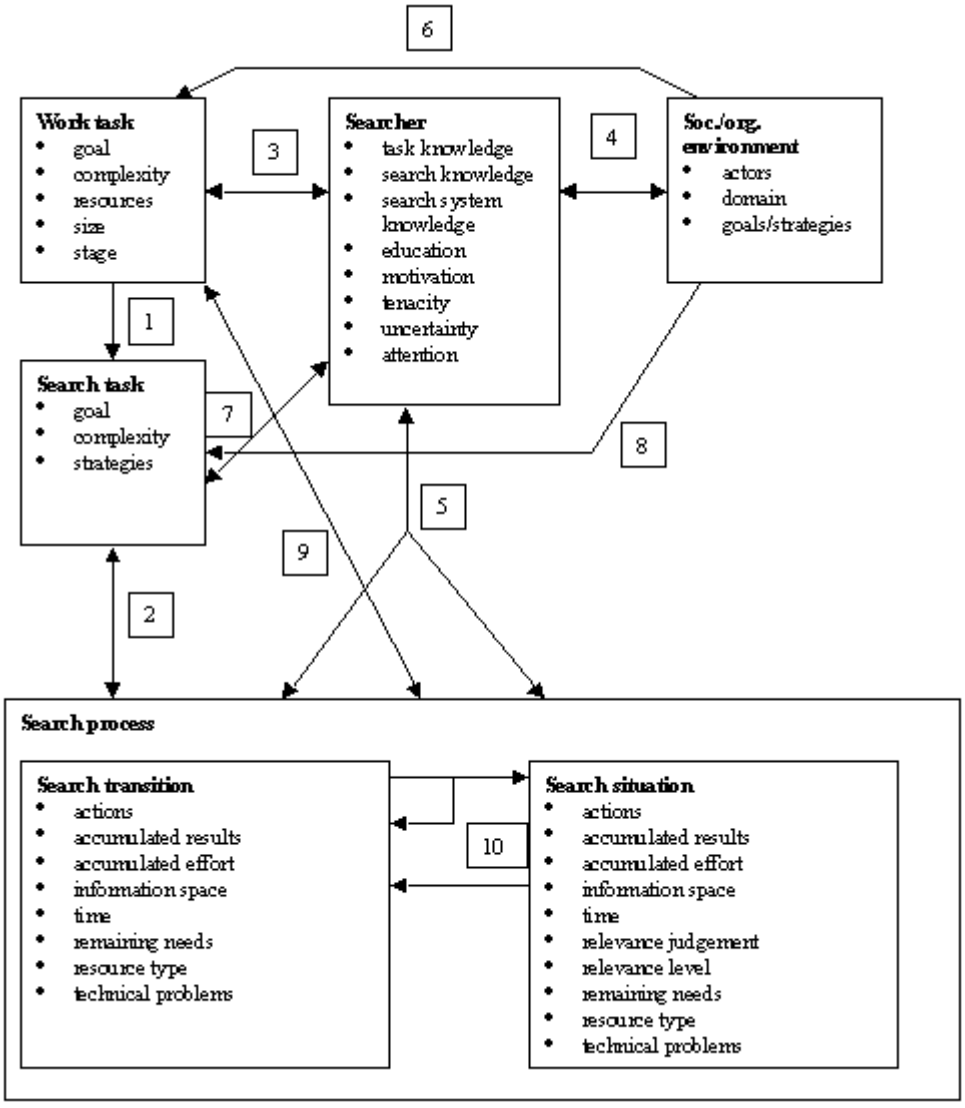

Figure 4.4 The conceptual framework - the domain of the method schema

Please note that the proposal of the following relationships should be treated as hypotheses, further research with different data sets is needed in order to find out whether these relationships are generally present. Some relationships have been treated in previous studies, e.g. the effect of various searcher attributes on information retrieval and seeking (the search process), while about other relationships, such as how work tasks affect the search process, less is known.

We shall examine one relationship at a time.

Work task – search task (1)

The content or goal of the search task will naturally be a product of the work task. The types of information needed (Järvelin, 1986; Byström & Järvelin, 1995) may differ, but the

search task will be generated by the work task – in special cases the search task and work task may be the same, for example for a reference librarian. A particularly complex work task may generate a multi-dimensional search task and the stage of the work task may indicate the search task goal.

Search task – search process (2)

The search task strategy is used to plan what resources to use and what actions to take in order to reach the goal of the search task, thus it clearly is a relationship between the search task and the search process. The search task complexity affects the time and effort invested in the search process.

The search process in turn affects the search task, by setting constraints on the strategies that are possible to execute. For example there may be resources that cannot be used as planned, or time spent may affect what ST goal(s) to pursue.

Work task – searcher (3)

There is a close interaction because the searcher constructs the (perceived) work task and the construction affects the behaviour.

The searcher's motivation is clearly influenced by the work task, since some tasks may be tiresome and routine and the searcher feels that they hinder her in performing more "exciting" tasks. Depending on the searcher's "personality", education, and experience, particularly complex tasks may be very motivating, while for others such tasks can trigger much uncertainty. Kuhlthau (1991; 1993) has identified how the work task stage influences the searcher's uncertainty.

On the other hand the searcher's knowledge about the work task obviously influences how she executes it.

Searcher – social/organisational environment (4)

The domain of the environment in which the searcher executes the work task influences her. It provides her with work task knowledge and the strategies and norms of the environment also may influence her search knowledge, i.e. they may provide rules on what kind of search systems she should use.

The searcher also influences the environment in which she takes part, for example by participating in the development of goals and strategies.

Searcher – search process (5)

There are many possible relationships between searcher and search process. Lack of attention may prevent her from identifying relevant resources or it may make her perform unnecessary actions. Great tenacity may benefit a searcher who has to wait for a long time in order for her action to be executed by a search engine. The latter example could also be used to illustrate how the search process influences the searcher, given that technical problems slowing down the process make the searcher less motivated.

Social/organisational environment – work task (6)

Work tasks may often be co-operative work and the actors in the searcher's environment may also be her colleagues on the same work task. Also the work task may be more or less specified by her nearest boss and clearly be influenced by the goals of the organisation.

Searcher – search task (7)

The searcher's search knowledge directly affects the search task strategies. In addition the strategies may be influenced by other searcher attributes such as motivation and attention, for example may an unmotivated and inattentive searcher formulate bad search task strategies.

Social/organisational environment – search task (8)

We believe that the goal or purpose of the search task only indirectly depends on the organisation, i.e. via the work task. Nevertheless the searcher's environment will influence the search task directly by guiding the search task strategies that are selected. This will happen partially indirectly, i.e. via the searcher's search knowledge. In addition actors in the environment may also give advice on how to attack the search task.

Work task – search process (9)

In addition to the indirect influence of work tasks, via search tasks, on search processes we believe that the work task also directly affects the process. Search processes are dynamic and during them the work task goal may be changed as a result of the results found, which in turn may invoke actions that were not originally planned. The stage of

the work task has been identified by Vakkari (2001) as influencing searching. In our case study we shall examine the work task – search process relationship.

Search transition– search situation [10]

Search situations and transitions are closely related. As we discussed above (subsection 4.5.5) they alternate during the search process. The attributes are to a large degree identical, i.e. the information space used is the same in situations and transitions although different parts of it may be activated. Time, effort, and results will accumulate during transitions as well as situations illustrating the tight connection between these two subcategories.

4.5.7 Summary of the method schema's domain

We have presented the main categories of our domain above by examining their attributes and exemplified how the categories can be related to one another. We have not presented every possible attribute that can be used to characterise dimensions of the categories, but have focused on those that were identified reviewing literature on information behaviour and verified as a result of our pilot study. While the relationships were used to show examples of how the categories could relate to one another, there are no doubt other inter-relationships between categories as well as intra-relationships, which we may see examples of in our case study.

Although we do think that these categories play a role, we do not know where, how, how often, with what impact or whether they are observable. We want to find out more about this and in Chapter 5 we will look at some of these issues in a limited setting.

When using the method schema to generate specific methods researchers will be able to identify other attributes that may be significant for their problems.

We shall now go on to focus on the second dimension of our method schema, the procedure.

4.6 Procedure of the method schema

The procedure of the method consists of a set of algorithms for data collection and data analysis. In Subsection 4.3.2 we examined the data collection algorithms[18] used in previous research on information behaviour, in Subsection 4.4.1 we presented the algorithms used in the pilot study, which were also used in the project.

We start by focusing on data collection in Subsection 4.6.1 and go through ways of analysing the data in Subsection 4.6.2.

4.6.1 Data collection procedures

The nature of one's research problem will often indicate whether one should choose a qualitative or quantitative approach for data collection. The terms qualitative and quantitative are most often used for denoting the data analysis, but if one wants to analyse information on a limited number of dimensions from a large number of searchers ("quantitative data set") then that would require other forms of data collection than if the analysis focuses on rich data sets from only a few searchers.

The method schema is developed to deal with research problems focusing on how work tasks influence Web information search processes. We believe that to use the method it is necessary to collect certain information about the searcher and her ways of dealing with work tasks and search tasks. This limits the collection algorithms.

Using a qualitative approach we recommend a combination of algorithms. In particular we believe it is fruitful to use interviews combined with some kind of screen capturing technique, and have the searchers talk aloud. Below we examine individual algorithms and the kinds of data that are derivable from them. First we discuss the use of real versus simulated work tasks.

Pia Borlund has proposed the use of simulated work task situations as an alternative way of evaluating interactive IR systems. Her approach is an alternative to the Cranfield model which "does not deal with dynamic information needs but treats information needs as a static concept entirely reflected by the user request and search statement" (Borlund, 2000, p. 5). She has compared how searchers behave when they try to solve search tasks related to their own problems (work tasks) and when they solve search tasks presented in the

[18] We here use "algorithm" to avoid confusion with respect to the use of the term "method". We use the term "procedure" as the common denominator for the set of data collection and data analysis algorithms proposed for the method schema.

context of a simulated work task situation. She concludes that it is possible to "substitute real information needs with simulated needs through the application of simulated work task situations" (Borlund, 2000, p. 140), but that it is necessary to tailor the situations to the searchers due to their motivation.

We find Borlund's findings interesting for two reasons. 1) She suggests a way to systematically investigate over a large scale how searchers behave solving similar work tasks in a controlled setting. 2) It may be easier to persuade searchers to be observed working with simulated work task than with their own work tasks. This implies, however, that the work task becomes a more "objective" entity, cf. discussions on perceived versus assigned work tasks (Byström, 1999).

If real work tasks are used then one should expect to have to invest time and effort in persuading searchers to assist in the study. If one is interested in large and quite complex work tasks it may be a good idea to use students who are supposed to work on theses of various kinds, or people working in organisations that have a high degree of project-oriented work. People working on small, and probably less complex, work tasks can be found in most organisations or in public places offering Web search facilities.

Interviews

The best way of obtaining information about the searcher's work task is probably by conducting an interview. In an interview situation the searcher may explain what is the purpose of her work task, at what stage she is in, what strategies she plan to use in performing the task etc. The interview may be designed to be open-ended and informal so as to let the searcher use her own words to explain the situation. The great advantage of the interview compared to questionnaires is that the researcher will be able to get answers to all questions, whereas in a questionnaire the interviewee decides what questions she wishes to answer.

A pre-search interview may provide information about work task, search task, the searcher and the searcher's environment. In a post-search interview it is also possible to get supplementary information about what happened during the search process. There are, however, data that may be just as well collected by using a questionnaire, for example demographic data. The questionnaire can, of course, be filled out in connection with the interview which would also solve the problem of "selective answering". Appendix 2 reproduces the interview guide used for our project.

Observation and video/audio recordings

Observation provides important data to use in order to analyse search processes. In addition the searcher should be instructed to talk aloud (Ericsson & Simon, 1996) during searching. Alternatively thinking-aloud could be used, but the latter requires more interpretation from the searcher and may prevent the natural flow of the search process. The observer should be in the background to take notes and remind the searcher to talk aloud, i.e. "to say out loud whatever they are saying silently to themselves" (Ericsson & Simon, 1996, p. 226).

It is recommended that the search performance is video recorded or recorded with the help of some kind of screen capturing facility. Her utterances should be recorded in a way that makes it possible to view the video and listen to the audio recordings simultaneously.

Diaries and journals

The searcher may also write diaries or journals recording the work task(s) performed and Web searching related to these. Then it is necessary to provide the searcher with a guide on what to make note of. The advantage of using a diary would be the ability to record short search sessions without the researcher having to be present. Often the most realistic scenario will be of a task performer who performs short Web information search processes several times a day rather than one long, continuous process.

The use of diaries and journals are however not suitable for identifying all the different attributes that influences during actual search processes. That would require the use of observation and/or video recording.

Document analysis

There may exist lots of written text that can be used to analyse work task-based WIS processes, for example the documents that were retrieved during the processes. The result of the task may be a written report or a thesis. The organisation in which the task takes place has most probably documented its activity thus providing information about, among other things, its goals and strategies. Depending on the research question, there may be other textual material to be used to supplement the data collected from the searcher.

Questionnaires

Questionnaires may be used to collect data from a large number of searchers. But they are also useable to supplement interview data in qualitative studies.

Demographic data are particularly suitable for collection via questionnaires, also information about the searcher's general information behaviour, such as what information systems she uses, the frequency of use, and her assessment of individual systems. Similar data can also be collected on different kinds of Web resources and the searcher's evaluation of the Web as a tool for work task performance. See Appendix 1 for the questionnaire used in our study.

Web transaction logs

It is possible to use transaction logs recorded at both the client and the server side of Web interaction.

Client side logs may typically provide data about the URLs visited, the technique used to access the URL and timestamps indicating for how long individual URLs were used.

Transaction logs recorded on the server side only, logically enough, record the use of URLs on that specific server. Such logs provide data about the IP-address of the client accessing the server, the client software in use, timestamps, and URLs of the pages that are accessed. Such logs are only of limited interest, if one wishes to analyse Web search processes that take place across different servers.

4.6.2 Data analysis

Data analysis can be said to take place at two different levels; first it is necessary to transform or code the raw data into analysable units, i.e. into categories, attributes, and relationships; then the coded data should be analysed in order to reveal how the different categories relate to one another.

Creating stories

In particular when having a qualitative approach, it can be difficult to separate between the collection and analysis of data. We recommend the creation of stories to describe WIS-processes; the procedure for making such stories was presented in Subsection 4.4.8, and we believe this procedure, "borrowed" from grounded theory (Strauss & Corbin, 1990) is fruitful. The stories are partly descriptive and partly interpretative, and we

recommend keeping the parts separated, for example by italicising the interpreted parts of the text.

Special kinds of software have been developed for transcription of field notes, Nudist (QSR N5, 2001) being probably the best known. We have no experience in using any such software but believe that they may be helpful in handling stories and ease analysis through the mark-up of categories and relationships.

An alternative way of computational processing of the stories, which is the way we have selected, is to use a word processing program and some kind of mark-up of all category occurrences and use, for example, arrows to indicate relationships. In this way it is possible to ease identification of the units of analysis.

It is of course possible to transcribe the data into other formats, for example by only focusing on the units of analysis without explicitly keeping the context in which they appear. When handling large amounts of data this may be more suitable, but it requires strict guidelines for classification. It is probably most suitable when examining the relationships between very few categories and/or attributes, for example what "resource types" are used in various "work task stages". This will typically be the case when combining observation/video recordings of many search processes and questionnaire data. We have not tested out this "algorithm" and therefore it should be considered tentative.

Content analysis

Content analysis (e.g. Holsti, 1969) can be used as a tool for both quantitative and qualitative analysis, depending on how the data to analyse were created. Traditionally it has been regarded as a tool for quantitative research, but we believe that it is also suitable to use this procedure to analyse stories created using the procedure referred to above. When a qualitative approach is selected, the content analysis starts already during story creation. The stories themselves are the result of describing what happens during WIS processes and interpreting why the process proceeds the way it does by relating it to its context, i.e. our knowledge about the searcher, her surroundings, work task(s), and search task(s). Analysing the stories by counting occurrences of words, sentences, and paragraphs constitutes a *quantification* of the data, and the analysis itself needs to follow the requirements of "objectivity, system, and generality" (Holsti, 1969, p. 3). This means that 1) it is necessary to develop clear rules for deciding what categories and attributes to use as well as what relationships to identify during story creation, 2) these rules should

specify when to include something as an instance of the category/attribute, and 3) it should be possible to compare the text/stories with one another. Appendix 3 reproduces our guide for classifying categories and attributes.

Relationships between categories and attributes that appear in the stories are the units to be counted in the content analysis.

The counting can be performed using a word processor's edit-find facility instructed to look for specific words combined with arrows, or one may manually scan through the text.

Qualitative content analysis takes place in two stages; in the story creation stage the interpretation of categories is performed combining all kinds of data collected; in the second stage the occurrences of relationships are counted. Quantitative analysis will consist of the researcher counting the occurrences of relationships, for example the number of journal articles, FAQs, and institutional resources that are used in different stages of a work task.

4.6.3 Summary of data collection and analysis

The research problem governs what kind of data to collect and the amount of data that is needed to explore it. Next a technique is needed in order to group the data into meaningful units of analysis. We propose the use of stories as exemplified in Subsection 4.4.8.

For analysing large amounts of data we believe it is possible to identify individual occurrences of a limited set of categories using observation and video recordings. Comparing observational data with data collected from questionnaires may then be used to analyse relationships between work task attributes and properties of the search process.

To analyse the data we suggest the use of content analysis, which can be used for analysing quantitatively as well as qualitatively generated data.

4.7 Evaluation of the method schema's domain (and procedure)

Our domain is represented as a conceptual framework in Subsection 4.5. It takes into consideration the work task, the searcher, the searcher's environment, the search task, and the search process. These represent important categories and were used to assess the

existing information behaviour model in Subsections 3.1.2 and 3.1.3. In this section we will first assess our framework using the formal criteria used in Subsection 3.1.1, of which C10 deals specifically with the procedure part of our method schema. Next we shall compare the categories and attributes covered by our method schema's domain with the content elements identified in the evaluation of the previous models.

In Table 4.6 we reproduce the table summing up the assessment of the existing models and adding a column for our own framework (SST-model).

	Wilson	Dervin	Ellis	Kuhlthau	Ingwersen	Belkin	Saracevic	Spink	SST
C1: Compatibility with previously well-tested knowledge	(+)	(+)	-	-	-	-	-	-	(+)
C2: Integration of formerly separate parts of knowledge	+	-	-	+	+	-	-	-	+
C3: Generalisation and explanation of lower abstraction level knowledge (observations, data) through higher level constructs	-	-	+	+	+	+	+	+	(+)
C4: Explanation of facts through systems of hypotheses which entail the facts	-	-	(+)	+	-	-	-	+	+
C5: Expanding knowledge by deducing new propositions based on selected starting points and collected information	-	-	-	-	+	-	-	+	+
C6: Pointing out fruitful problems	-	+	+	+	+	+	+	+	+
C7: Proposing the collection of data which nobody would understand to collect without the theory	-	+	-	+	+	-	-	-	-
C8: Proposing new lines of research	+	+	+	+	+	-	-	-	(+)
C9 Representing or modelling the objects (and relationships) of the relevant part of reality instead of plain summarisation of data	-	-	-	+	-	-	-	-	+
C10: Suggesting tools for producing new data	-	+	+	+	-	-	-	+	+

Table 4.6 Summary of previous models and the SST method schema's ability to meet the formal criteria of Bunge (1967)

As discussed above there is little "well-tested knowledge" to be compatible with in the field of information behaviour. Assessing previous models' ability to capture what we believe are essential categories influencing WIS processes, we have indirectly shown how our framework relates to past works, thus the partial value. It does integrate elements from previous models, combining categories derived from information seeking as well as information searching. C2 could therefore be said to be satisfied.

Criterion 3 has to do with the generalisation and explanation of observations/data into higher level constructs, which in our case means the generation of categories representing observable and interpretable features of WIS processes as well as attributes representing characteristics of the searcher, her work task etc. The SST model provides a basis for such generalisation and explanation. If the method schema is justified we may state that this criterion is satisfied and thus we cannot at this stage claim that it satisfies C3, but we indicate its potential success by giving it a partial score. We believe that the SST model satisfies C4; it explicates reasons for WIS processes by stating relationships between the categories and exemplifying how individual attributes relate to one another. C5 is also met, since our method schema's domain is based on previous model's findings and uses such findings to point out new relationships.

The framework suggests fruitful research problems (C6); one may, for example, examine the direct influences between work tasks and WIS processes, i.e. those influences that are independent of the search task.

We do not, however, claim that the framework proposes collection of data that otherwise would not have been collected. Criterion 7 demands, in our opinion, that the data to be collected should be original within the discipline while our framework represents more of an accumulation and combination of previous research. The SST model proposes new ways to look at the data, but no novel data as such. The method schema is partly derived inductively, partly based on knowledge contained in previous works, and it clearly relates to the cognitive IR interaction model (Ingwersen 1992; 1996), as well as Belkin and colleagues' (Belkin, Marchetti & Cool, 1993) episodic ISS-schema. It does go a step further in operationalising the relationship between information seeking and information searching/user-centred IR than Ingwersens's model. We may therefore state that it partly satisfies C8.

The last two formal criteria have to do with mapping of reality and we believe our framework clearly satisfies C9 – modelling the objects (and relationships) of the relevant part of reality. C10 has to do with tools for data production (collection and analysis).

Since this has been specified by our method schema's procedure our work clearly satisfies it.

In Table 4.7 we present the content elements treated by all the models that were evaluated in Subsections 3.1.2 and 3.1.3 and add the categories and attributes included in our method schema's domain.

When comparing the content elements of our method schema's domain with the existing models of information behaviour, we see that our method schema suggests a larger number of categories affecting information behaviour than previous models. Notice that the knowledge attribute of the searcher is developed in more detail in the SST method schema. Ingwersen's 'model' and Saracevic's 'knowledge' are thus (partially) credited with covering work task knowledge as well as search knowledge. Belkin is "credited" for separating interaction in "information resources" (situation) and "meta-information resources" (transitions).

The content-based and formal criteria indicate that our schema should be suitable for generating methods to reveal what takes place during WIS processes.

The explicit focus on analysis of *processes* is an important feature of the SST method schema. In particular its ability to allow detection of how external categories such as work task, search task, and searcher affects the process.

	Wilson	Dervin	Ellis	Kuhlthau	Ingwersen	Belkin	Saracevic	Spink	SST
Work task	-	-	-	-	(+)	-	(+)	-	+
• stage	-	-	+	+	-	-	-	-	+
• goal	-	-	-	-	-	-	-	-	+
• complexity	-	-	-	-	-	-	-	-	+
• resources	-	-	-	-	-	-	-	-	+
• size	-	-	-	-	-	-	-	-	+
Search task	-	-	-	-	(+)	-	-	-	+
• information need	+	(+)	-	-	(+)	-	-	-	-
• goal	-	-	-	-	-	-	-	-	+
• complexity	-	-	-	-	-	-	-	-	+
• strategies		-	-	-	-	-	-	+	+
Searcher	+	-	-	-	+	-	+	-	+
• uncertainty	-	-	-	+	+	-	-	-	+
• work task	-	-	-	-	+	-	+	-	(+)
• problem/goal/ intent	-	-	-	-	+	-	+	-	-
• information need	-	+	-	-	+	-	-	-	-
• model/knowledge	-	-	-	-	(+)	-	+	-	
• task knowledge	-	-	-	-	(+)	-	(+)	-	+
• search knowledge	-	-	-	-	(+)	-	(+)	-	+
• search system knowledge	-	-	-	-	-	-	-	-	+
• education	-	-	-	-	-	-	-	-	+
• motivation	-	-	-	-	-	-	-	-	+
• tenacity	-	-	-	-	-	-	-	-	+
• attention	-	-	-	-	-	-	-	-	+
Environment	+	-	-	-	+	-	+	-	+
• domain	-	-	-	-	+	-	-	-	+
• strategies/goals	-	-	-	-	+	-	-	-	+
• tasks	-	-	-	-	+	-	-	-	-
• preferences	-	-	-	-	+	-	-	-	-
• actors	-	-	-	-	-	-	-	-	+
Search process/ interaction	-	-	-	-	+	+	+	+	+
• *transition*	-	-	-	-	-	(+)	-	-	+
• *situation*	-	-	-	-	-	(+)	-	-	+
• method/query/ move	-	-	-	-	+	+	+	+	(+)
• goal	-	-	-	-	-	+	-	-	-
• mode	-	-	-	-	-	+	-	-	-
• resource	-	-	-	-	-	+	-	-	+
• action	-	-	-	-	(+)	(+)	(+)	(+)	+
• accumulated results	-	-	-	-	-	-	-	-	+
• accumulated effort	-	-	-	-	-	-	-	-	+
• information space	-	-	-	-	-	-	-	-	+
• time	-	-	-	-	-	-	-	-	+
• remaining needs	-	-	-	-	-	-	-	-	+
• technical problems	-	-	-	-	-	-	-	-	+
• relevance judgement	-	-	-	-	-	-	-	-	+
• relevance level	-	-	-	-	-	-	-	-	+
Search system	-	-	-	-	+	-	+	-	(+)
• search language	-	-	-	-	+	-	-	-	-
• indexing rules	-	-	-	-	+	-	-	-	-
• database structure	-	-	-	-	+	-	-	-	-

Table 4.7 Content elements in existing models and the SST method schema

4.8 Summary

Our method schema was developed using a grounded theory approach. This has been particularly useful in that we have tried to have an exploratory view focusing on how different categories affect the WIS process at the same time as aiming at what the categories are from a WIS point of view.

We deliberately left the definitions of the categories open in order to see how new data fit into the schema. This provides a flexible tool which is open for the addition of new attributes and categories and which suggests relevant strategies for conceptualising WIS process problems.

In Chapter 5 we present the use of a method to examine work task-based WIS processes performed by students in LIS. This constitutes the justification part of our method schema.

5 An empirical study of Web users – justification of our method schema

The present chapter examines an empirical study of students using the World Wide Web as a tool for the solving of real work tasks. The empirical study provides the *justification* part of our method schema. In order to analyse their WIS processes and the factors affecting them we have derived a method from our method schema. In this chapter we first present our empirical research questions, then we examine how we have collected our data, thereafter we consider the rules of classification for the analysis of the data, before we present the results of the analysis.

5.1 Demands to a justification

We need to justify our method schema's applicability as a tool for understanding the interplay of different factors during WIS processes. This means that we should provide arguments that are good enough to convince the research field that the SST model is worth using for identifying interplay between work tasks and other factors that are thought to play important roles during WIS processes.

Our justification has two parts. The major argument is a case study where a set of empirical research questions is examined using data from real WIS processes. Justification can be achieved partly by using a method derived from the schema to answer the research questions in a convincing way. The second part of our justification consists of an evaluation of the schema after the formal criteria that were used to evaluate existing models in Chapter 3. For such an evaluation we refer the reader to the foregoing chapter, Subsection 4.8.

5.2 Empirical research questions

Many modern approaches within information seeking and searching (or retrieval) emphasise the importance of tasks. Although it may be unclear whether they focus on work tasks or search tasks (Vakkari, 2003), it is an even bigger problem that their conceptualisation remains on too general a level to reveal how work tasks and search tasks affect the search process. In our method schema we have developed a conceptual framework that takes into consideration both kinds of tasks and we treat them on a level that prepares for the analysis of empirical data. It is therefore natural for us to try out our method schema by proposing empirical research questions that have to do with work tasks and search tasks.

It is theoretically and methodologically important to find whether the work task directly affects the search process or if it is only indirectly through the search task. If the work task only influences through the search task then we can simplify study settings, for example by separately studying the relation work task – search task from the relation search task – search process. If there is no, or hardly any direct work task – search process impact, what do then search tasks consist of other than the search topic; how much does the searcher bring with herself? In our method schema we have hypothesised that there is a direct effect of work tasks on WIS processes, thus this should be one focus of our empirical research. The identification of the interplay of work task and search task may provide better definitions of these concepts for the LIS field.

In addition we believe it is relevant and fruitful to see how searchers' knowledge on work tasks and search tasks affects the search processes. Previous research has been able to show that search and topic (work task) expertise influence search tactics (e.g. Hsieh-Yee, 1993). With our method we may see whether such knowledge types affect other characteristics of the WIS process and whether the effects are positive or negative.

The following research questions were raised:

- How do work task goals influence web information search processes?
- How do work task stages influence the web information search processes?
- How do search task strategies influence the web information search processes?
- How does search knowledge influence the web information search process?
- How does work task knowledge influence the web information search process?

In addition we wished to examine differences and similarities of situations and transitions in order to see whether this specification of WIS processes is fruitful.

We have used a limited set of data and for that reason the questions cannot, on a generalisable basis, be answered using the data we have. The questions can, however, be used to explore the relationships of the categories and attributes to show that the method schema and the method we have derived from it works in analysing work task and search task effects on Web information search processes at the level of situations and transitions. The data collection and analysis is illustrative of how the method schema can be used, but we can only give answers concerning the influence of the categories for the processes we have examined.

5.3 Collecting the data

In Subsection 4.4 we presented our pilot study including the data collected for our project. We asked all final year students in library and information science at Oslo University College to participate in the study. LIS students were selected because at the time of data collection (1997) these students were fairly advanced Web users compared to students at other faculties at the College; they received training in Web searching and were also offered a course in Web publishing. The researcher also had physical closeness to these students, since he worked on the faculty, thus facilitating flexibility with respect to the set up and implementation of observation of the students. Below we give a short introduction to the characteristics of the students.

The data consist of four different types of material: results of a questionnaire; interviews; observation notes; and video recordings of search sessions. In addition the completed theses have been used to gather supplementary data. The data used in the pilot study consisted of three selected search sessions. These three sessions and six additional sessions have all been used in our empirical study.

The fact that the data used for developing the method schema was also used in the empirical study in order to justify the schema raises the serious question of circular justification. We will claim that the use of the data from the three sessions served two different purposes. In the pilot study we used the data to develop categories and attributes while in the empirical study we focus on the relationships between selected sets of attributes. The relationships proposed as a result of the pilot study in Subsection 4.5.6 was to a large extent based on deductions by the researcher. In the empirical study we propose a set of research problems that focus on specific relationships and analyse the data in order to examine systematic occurrences of such relationships. Thus the data are used differently. We believe that by using all 9 sessions in the empirical study we gain more in terms of showing how the schema is applicable as a tool for this sort of analysis than what we would earn from only using 6 sessions. Even when we take into account that the latter approach would spare us from the circular argumentation problem. We prioritise to have a larger data set in which to perform our empirical analysis. The issue of circular justification may perhaps be a small problem, but not a big one.

In qualitative study, it is typical to develop the analysis schema by repeatedly analysing larger and larger subsets of data before analysing the whole data set. This can be seen in contrast to the procedure in typical IR experiments where a learning collection is used to learn parameters for some method, and a separate test collection is used to show that the parameters deliver the expected performance.

5.3.1 Student characteristics

Oslo University College's course of education in librarianship takes three years. The first year has a common content for all students, including subjects concerning "document and users" on the one hand and "information storage and retrieval" on the other. In the second year some subjects are optional and in the third year students choose whether they want to specialise on the one hand in "Information retrieval and storage" or "Literature and society", the latter includes subjects like e.g. "fiction studies", "Norwegian cultural history" and "contemporary literature". On the other hand in "Library organisation and management" the students choose to specialise towards either public libraries or research and special libraries. There is a strong interdependence between the specialities chosen by the students and a large majority of students choosing "Literature and society" also chooses to specialise in public library management. Correspondingly most students who choose management of research and special libraries also choose "Information retrieval and storage". Besides the special subjects students follow courses in methodological subjects such as mathematics, linguistics and computing.

For their thesis work students usually choose a subject within their speciality. That year (1997) the students had also been offered a course in organising Web resources directed towards thesis writing. Quite a few students had chosen to participate in this course and this has resulted in a new kind of thesis writing. The web organising course, for the main part, has resulted in theses in the form of Web sites for various organisations. A few students have also chosen to use the Web as a means for other kinds of theses, for example constructing a thesaurus or comparing web genres with their corresponding paper genre. These theses however have their main focus on the other subjects.

5.3.2 The questionnaire

All students were asked to complete a questionnaire about demographic facts (age, gender, and education), computing skills, Internet use, and information behaviour related to their thesis. The questionnaire is reproduced in Appendix 1.

Along with the questionnaire came a letter signed by the leader of the national research program on library and information science and the dean at the Faculty of Journalism, Library and Information science, which urged the students to participate in the study. In addition a form was enclosed where the students could indicate if they were interested in being observed during thesis-related Web searching. The students were not given any

financial compensation, but were told that among the students being observed we would draw three winners in a book lottery.

To collect the data it was necessary to distribute many reminders in various formats. We co-operated with lecturers to reach all students in a classroom situation, four such classroom situations being needed to reach the students. The students were given a short introductory lecture on the purpose of the study and were asked to return the questionnaire as well as to volunteer as candidates for observation. The questionnaires were handed out at the end of each lecture and we put up posters to remind the students to return the questionnaires.

Unfortunately only a relatively small number of questionnaires were returned, which prompted us to remind the students by handing out a new copy of the questionnaires at the end of lectures. In addition we put up new posters in central areas of the institution visited by the students. The limited success of this reminder made it necessary for us to send out personal copies of the questionnaires to all students who had not returned it after a month.

The major findings from this part of the study were reported in Pharo (2000). The most important contribution of the questionnaire-generated data has, however, been to supplement the observed Web sessions with knowledge about the searchers' backgrounds; their education, search knowledge, Internet knowledge etc. We choose to present some background material about the work tasks (the theses) performed by the students and how they view the Web as a source of information for solving their tasks.

In all 55 students returned the questionnaire. Two students only returned a small part of it and have been excluded from this survey. This means 50 % of all third year students responded, a surprisingly low figure taking into account the fact that the survey was presented for the students on several occasions and that they were reminded to return the questionnaire several times in several ways. The fact that the participants were self-selected obviously leads to a bias in the material. This is, however, not a problem in our case since we do not wish to generalise among all students, but rather to study real work task-based WIS processes. It may be that the students not answering the questionnaire saw themselves as non-users of the Web and chose not to participate because the survey is part of a project investigating Web use. It may also be that the focus on "information sources" did not seem relevant for these students, in the sense that they did not intend to use any external sources or did not consider their fellow students or lecturers "information sources", but rather saw them as "co-operation partners". There are,

however, quite a few students working on projects definitively requiring web use who have not answered the questionnaire and among the received answers there is a substantial portion of non-web users. It is therefore hard to point at this as a feature of the students not returning the questionnaire. It would be interesting to learn whether there is a terminology bias concerning "information source" in a population in which studies of "information storage and retrieval" form a substantial part.

The 53 questionnaires analysed represent students working on 38 different theses, i.e. approx. 57 % of this academic year's 67 different theses were represented in the material. We have categorised the theses according to what main subject they belong to: "Literature and society" (LISO); "Information storage and retrieval" (INFO); " Library organisation and management"[19] (ADM); and "Web publishing" (WEB).

Subject category	Number of projects	Number of students
LISO	12 (32 %)	14 (26 %)
INFO	11 (29 %)	13 (25 %)
ADM	8 (21 %)	11 (21 %)
WEB	7 (18 %)	15 (28 %)

Table 5.1 Distribution of projects and students per category

Table 5.1 shows the distribution of all students by project or thesis type. There is a difference in the distribution of students per thesis type, since on average 2.14 students work on a Web project, while there are only 1.17 students per LISO thesis. INFO and ADM theses on average are performed by 1.18 and 1.38 students respectively. The difference in the number of students per project may be due to several reasons, but we shall see that most WEB theses are the result of external assignments, which may be of a kind and size that require a group of students to co-operate to ensure their completion.

To the question "How did you select the topic of your thesis" the students gave the following reasons:

Initiation method 1: Defining it themselves (own)
Initiation method 2: Lecturer's proposal (internal)
Initiation method 3: Assignment from external principal (external)

[19] We have not differentiated between public and research library management, since these theses often focus on problems of interest to both library categories.

Initiation method	ADM	INFO	LISO	WEB	*Total*	
Own	4	2	10	1	*17*	*32%*
Internal	4	7	2	2	*15*	*28%*
External	3	6	2	10	*21*	*40%*
Total	*11*	*15*	*14*	*13*	*53*	*100%*

Table 5.2 How the projects were generated

Table 5.2 shows how the different projects were generated. Very interesting are the differences in how LISO and WEB projects are initiated. While almost all of the students (10 students) performing literary analysis and other LISO projects have defined their own projects and very few are proposed by lecturers (2) or are external assignments (2), we see the opposite trend when examining the WEB projects; 10 projects are initiated by external principals, 2 by lecturers at the faculty, and only one WEB project is initiated by a student. It was noted above when discussing the number of students per project that LISO students probably preferred to work on problems of a more individualistic nature and this is confirmed by the results presented here. LISO projects are without doubt more strongly founded in the students' basic interests, i.e. a general interest in fiction and reading acquired long before starting their education to become a librarian, and probably the strongest reason for their wish to become one in the first place.

Above we hypothesised that Web projects might demand more students, since the expectations to these are higher and demand more diverse expertise, for example more students with different specialities or more work power. The fact that most WEB projects are initiated from outside the students' environment is probably due to the marketing of students' web publishing skills towards (non-commercial) institutions which feel an emerging need to expose themselves in the new medium. Whether this high external demand continues in the future will be interesting to see. It is also interesting to note that few students themselves initiate INFO projects. Quite a few of these are also the results of external initiative (e.g. creating indexes to journals), but lecturers obviously play an important role with respect to projects concerned with information storage and retrieval. The fact that this is a subject new for most students when starting their education is probably a reasonable explanation for the students turning to their lecturers for good ideas. Administration and management is a more general subject area than INFO, thus explaining why the origins of these projects are more evenly distributed.

When evaluating the information sources the students had used and planned to use in their thesis work, we had asked them to mark the importance of the source on a Likert scale.

In all, 5 different categories of sources/channels were presented; persons; library and information services; printed material; Internet services; and the obvious "Other" category. Apart from World Wide Web, Internet services included FTP, electronic mailing lists, and Usenet newsgroups. Students were asked to evaluate the sources/channels on a 1-7 scale, grade 1 meaning that the information service (IS) was not important and grade 7 meaning that the service would be or had been of great importance.

To compare the importance of the different information services a simple measurement is introduced; the Information Service Importance Rate (ISIR). The ISIR is measured by dividing the sum of all grades on the number of grades, i.e. the average grade, and can be used to quantify the importance of individual information services for different categories of users.

ISIR = sum of grades/number of grades

Figure 5.1 The Information Service Importance Rate - ISIR

The ISIR may be used to measure a single student's evaluation of information services' importance or all students evaluation of a single IS. We have used it to measure the students' evaluation of a selection of information sources and services.

Table 5.3 shows how the Web and other potential information sources are valued by the different categories of students. We see that the students' lecturers are considered to be the most important sources of information. It is also interesting to note that journal articles are considered to be less important for these students than human resources. Possibly these students are at an academic level where textbooks are more important sources of information than journal articles. There are several reasons why students working on WEB projects consider the Web to be such an important IS compared with the other students. One important factor is probably that texts concerning the Web are first published on the Web and that the goals of WEB students' projects are to construct Web sites, i.e. their work context is the Web itself. These students are thus very oriented towards the Web and probably more aware of the conglomerate of ISs available there. It is interesting to see that INFO students rate Web higher than the other groups (except for the WEB group), but it is hard to say whether this is because they are more familiar with the medium or there is more potential information on the Web aimed at this group.

	Lecturers	Fellow students	External person	Books	Journal articles	Usenet news	Web
ADM	2.67	1.67	5.33	1.8	3.67	2	1.64
INFO	5.23	2.69	2.92	5.67	1.72	1.33	2.2
LISO	4.11	2.89	1.83	4	2.67	2.44	1.43
WEB	4.23	4.97	4.25	1.46	1.38	5	4.85
Total	4.26	3.3	3.19	3.5	2.23	2.82	2.83

Table 5.3 Students' evaluation of a selection of sources' importance as information services for their project work

The questionnaire results revealed that one particular group of students was particularly interested in using the Web as a tool in solving their work tasks – those students working on Web related projects. Of the 14 students who originally volunteered to be observed during Web searching, 10 (!) worked on Web publishing projects. Thus the students that we ended up observing represented a limited set of work tasks and had quite high expectations of what to find on the Web in order to help them solve their tasks. In the end only 10 students contacted us in order to be observed and of these 6 worked on Web projects. However, the only session performed by two of these students was not recorded due to technical problems. We shall return to the characteristics of the recorded sessions below.

Examining the students specifically choosing to participate in the study, we found that they considered their Web experience to be medium; 3.8 on a 1-7 Likert scale, which was slightly higher than the average score for all students (3.37). Their Web use was divided between being recreational and informational. Most students had been fairly new to the Internet when they started their education in LIS. Their experience with computers they also regarded to be average (4.15), which was also slightly higher than the average score for all students (4.0). We can state from this that the students choosing to participate in our study did not constitute a group of experts, but rather believed that they were in the middle range with respect to computer and Web experience.

5.3.3 Pre-search interviews

Before each search session the students being observed were asked about a set of issues concerning the forthcoming session and the current status of their project.

We designed the interviews to be open ended and informal, which means that the interviewees used their own words in responding to the questions, which were not put in a pre-designed form. The interview guide (see Appendix 2) centred around the following issues: the theme of the thesis, the current status of the work task and progress since last session, the feelings of the searchers concerning the work task, and the goal of the forthcoming session. We had developed different angles to focus on the different issues, which were used when it was felt necessary.

The interviews took place with all the searchers who were supposed to participate in the session. In addition we conducted one interview with four students who worked on the same thesis and were planning to be observed the same day. The pre-session interviews were different in length, lasting from only a few minutes to a quarter of an hour, after which the sessions were started. Interviews were conducted in the same room as the Web search sessions were performed.

We took notes from the interviews and have included our interpretations of the work task goals and stages as well as search task goals, complexity, and strategies in the stories generated for each search session.

5.3.4 Observation and video recording

The students were instructed to contact the researcher whenever they wanted to use the Web as part of their thesis work. To make the Web sessions as natural as possible the students used their own accounts to log into the local network and used their own version of the client software (Netscape 3.03) with their own bookmark files and configuration. The only change made in their set-up was to increase the font size in order to enhance reading when studying the videotapes. The users were also allowed to work together if they wished to do so.

The sessions were recorded in an undisturbed room and with the help of a GrandArt video converter, which is a device that converts computer screen signals to video format. A microphone was connected to the converter and the users were asked to talk aloud (Ericsson & Simon, 1996) during the sessions, to make it easier to identify their intentions and goals. Their utterances were recorded simultaneously on the videotape.

During observation the observer made notes of what appeared to be significant events. In the previous chapter (Subsection 4.4.4) we showed how such notes were used in the first

step of coding the data for the pilot study, pointing out important characteristics of the search process-category.

Nine Web sessions, in total 12 1/2 hours of data, were analysed. Of the nine sessions three were performed by two students working together. The nine sessions were related to three different theses. Seven sessions dealt with a WEB project, while the remaining two were connected to two different projects; one aimed at analysing the structure and content of web-based newspapers; the other was a bio-bibliography of a Norwegian author. The latter two were both classified as INFO projects.

The video recordings constitute the most important data for creation of stories. We transcribed each session into the format described in Subsection 4.4.6, the point of departure being the actions performed. In addition we recorded the resource used, utterances by the searcher, and time. Table 4.4 in Subsection 4.4.6 shows an excerpt from one of the transcribed sessions.

5.3.5 Post-search interviews

After every session we asked the students some follow-up questions. These interviews were often very short due to the students' needs to fulfil other obligations and at most they lasted five minutes.

In the post-search interviews we aimed at getting an evaluation of how the students felt the session had helped them with their thesis (work task) and whether they could have used alternative sources for the information need concerned (search task). We also wanted to know how they planned to continue their thesis work.

The information gathered in the post-search interviews has been of limited value, although to a certain degree the interviews have helped us in establishing the purpose of individual sessions. An alternative approach could have been to go through the session with the searcher to let her comment what happened, unfortunately this was not possible due to limitations in time. In addition we believe that most searchers were quite exhausted at the end of the session, which would have made it difficult to concentrate on such a task.

5.4 Creating stories

We chose to create our stories founded on the transcribed video recordings and supplemented with observation notes, interviews, and questionnaire data. Stories are

based around the transitions and situations taking place in the search process. We can categorise situations/transitions by identifying instances in the transcription notes when the searcher switches between resources and/or changes to a new search strategy.

The stories can be said to be a second-order transcription of the video recordings. They are partly descriptive and partly contain interpretation of the actions, resource types in use and utterances made by the searcher using the categories and attributes of our method as a classification scheme. The procedure presented in Subsection 4.4.8 has been used to develop the stories.

5.4.1 Validity of story creation

Before presenting the guidelines for classifying categories and attributes it is necessary to express the following reservations.

- the stories are the researcher's interpretation of the Web information search processes;
- the stories are faithful to the data that were collected, i.e. the interviews, observation notes, and video transcripts;
- the stories are considered to represent plausible interpretations;
- the stories go as far in representing the data as possible without becoming speculative;
- there is a lot of interplay between various categories that goes unnoticed in the stories and remains unobserved in the heads of the searchers;
- it would be difficult to grasp more of the searchers' thoughts while still keeping the search process fairly natural.

The stories represent as much as we can find out about the effects of work tasks and other categories on the search process. There are, however, also undiscovered effects. Nevertheless, if it shows that work tasks directly effect the search processes, there is a proof of its existence. If nothing should be found, the effect is not disproved, just not captured.

We have chosen to represent the search sessions as *stories*, which are created by combining various kinds of raw data. Research on complex phenomena by using a combination of several kinds of data focusing on different aspects of the phenomenon is called triangulation. The stories could be said to be the condensation of data collected using triangulation. We have used data stemming from questionnaires, interviews conducted before and after search sessions, observation notes, video recordings of sessions with searchers talking aloud, and the final products of the work tasks, i.e. the

theses submitted by the searchers. The search session recordings were obviously the core data for producing our stories, and by comparing the video transcripts and the stories (see Appendices 4 and 5) the reader will note that the stories to a large degree are based on these transcripts. Indeed the descriptive parts of the stories are directly derivable from the video transcripts. We believe that to combine the various kinds of data into stories by supplementing the descriptive elements of the video transcripts by knowledge about searchers work tasks, search tasks, and personal attributes such as their search knowledge and work task knowledge gave an "added value effect" that would be harder to find if each kind of data should be treated on its own. The story format made it possible to relate, for example, information provided by the searchers about their work tasks to the actions executed during the sessions, or the web model ("information space") activated during the sessions. Each event in the stories, and each interpretation in them, can be traced back to one or more data sets. We have no data on the searchers' thoughts, just their actions and words (interview data, comments made during the process, theses). It is difficult to get closer to purposeful (meaningful) process of each searcher? For these reasons we believe our story-telling approach is valid.

To secure consistency of coding a coding handbook was developed (see Appendix 3). See also Subsection 4.4.8 on story creation.

To secure the reliability of the data we let the stories rest for three months before performing a second classification of a limited set of the stories. 20 per cent of the stories were reclassified, representing parts of the stories in the beginning, end, and the middle of the search processes. The reclassification was designed so that the unclassified parts of the stories were gone through to make certain that we were able to take into consideration the accumulation of results and effort.

Our reliability testing showed a consistency of between 65 and 73 %, which indicates that the classification of the sessions is reasonably consistent. One should also take into consideration the observations referred to by Fidel (1993, p. 232) that classification or coding consistency seems to decrease as the complexity of the phenomena under study increases. The stories are quite complex and thus we find the consistency rate to be satisfactory.

5.4.2 How to classify categories and attributes

We have developed a set of rules facilitating the use of the categories and attributes of our conceptual framework to classify the stories. These rules are reproduced in Appendix 3.

The rules partially state the content of some attributes in those instances where those attributes are the same for all searchers. In most cases, however, the rules state how to identify the categories and attributes. We do not propose a final set of possible values that can be given to each of the categories. The main purpose is to identify whether the individual category/attribute influences other categories and attributes during the WIS processes.

In the classification rules we have explicated whenever an attribute should be categorised, and for some of the attributes characterising categories 1-4 we have stated that they should be coded if, but only if, we can observe it having an effect. Other attributes should be coded whenever they are identified.

5.4.3 How to classify relationships?

Not all identifications of an attribute reveal a corresponding relationship of that attribute to at least one other attribute. Such relationships may exist, but we may not have the proper information to identify them.

In addition the relationships *may* be classified as having positive or negative effects. This, in particular, is the case for relationships generated by attributes that in some way characterise the searcher, i.e. all 2.1 – 2.8 attributes (see Appendix 3).

Positive and negative values

In addition, attributes can assume positive or negative values. The searcher can have lack of task knowledge, education, motivation, tenacity, uncertainty, attention, and/or search knowledge. This may complicate our categorisation of relationships, since for each occurrence of task knowledge four different relationships are possible:

Task knowledge +++> attribute x (positive effect)
Task knowledge ---> attribute x (negative effect)
Lack of task knowledge +++> attribute x (positive effect)
Lack of task knowledge ---> attribute x (negative effect)

We propose that the following notation should be used:

1. "attribute x + → attribute y" is used to state a positive effect as a result of attribute x on attribute y
2. " attribute x - → attribute y" is used to state a negative effect as a result of *lack of* attribute x on attribute y
3. " attribute x + - → attribute y" is used to state a negative effect as a result of attribute x on attribute y
4. " attribute x - + → attribute y" is used to state a positive effect as a result of *lack of* attribute x on attribute y

For the remaining attributes, i.e. those being attributes of the searcher, we use "+ →" to mark positive effects and "- →" to note negative effects.

The assigning of negative and positive values may cause some semantically related problems. "Uncertainty" has a negative connotation and thus it can be confusing to state that one factor negatively affects uncertainty, since it can also be interpreted as meaning a decline in uncertainty. We do however choose to assign positive and negative values to the *relationship* from the *viewpoint of the task performer*, i.e. if some attribute is identified as having a negative effect on the searcher's performance. Examples of such negative effects are: an increase in time spent; an increase in effort invested; something causing a misrepresentation in the information space; whatever causes a technical problem etc. If some relationship is detected that hinders the searcher's efficiency during searching then it should be marked as a negative effect, and vice versa; it should be identified when positive effects occur.

5.4.4 Coding

In the stories the classification codes should be spelt out with number and name to ease readability. If a relationship is identified, the relationship notation should be put next to the attribute causing the effect, and to ease readability each attribute should be put where it appears in the transition/situation. The codes should be enclosed by parentheses. All transitions and situations have been represented in a descriptive form, but in some cases it has been necessary to add an interpretation in order to clarify what happened during the transition/situation. These interpretations, which are formatted with Italics, should also be coded but the classifier should be aware of redundancy, i.e. try to avoid coding the same occurrence of an attribute/relationship more than once.

5.4.5 Interrelationships

In some cases two attributes can have a two-way relationship, which in philosophical terms we may call a hermeneutic relationship or circle. This is particularly apparent when we look at how a searcher's information space can be changed as the result of interaction. The best example is to look at how a searcher interacts with an unfamiliar resource type. Does her existing information space help her in interpreting how to use the resource or does the resource change her information space? We choose to use the following notation in such cases:

5.1.4 (5.2.4) information space ←→ 5.1.7 (5.2.9) resource type

5.4.6 The action – resource type relationship

A special case of relationship exists between 5.1.1(5.2.1) action and 5.1.7(5.2.9) resource type. Since an action necessarily has to be performed within or on a resource, we have decided to denote this relation with the use of the hyphen unless it can be split up in the story. The following sentence exemplifies the relationship: "She selects a link from Yahoo's subcategory on computer games". This sentence can be coded in the following two equal ways:

1. She selects a link (5.1.1 action) from Yahoo's subcategory on computer games (5.1.7 resource type)
2. She selects a link from Yahoo's subcategory on computer games (5.1.1 action – 5.1.7 resource type)

5.4.7 Some examples

We here present some examples showing how to use the model. The following excerpts are taken from real stories.

> They have heard that SF Gate is a good example of a successfully implemented web newspaper (*2.1 WT knowledge* + → *2.5 motivation*). In S9_S6 they start at the site's start page (*5.2.9 resource type*) where they follow the link (*5.2.1 action*) to its subject index (*5.2.9 resource type*) in which they spend some time discussing what to look for. They eventually choose to look at the paper's news page (*5.2.9 resource type*) and comment that it (the newspaper generally) is so different from a Norwegian paper's solutions (*5.2.9 resource type*) that they consider it not relevant for their purpose (*1.1 WT goal* → *5.2.7 relevance level* → *5.2.6 relevance judgement*). They also take a look at a page dealing with Internet issues, which they

> do not find relevant (*5.2.6 relevance judgement*). Finally they go to the current digital edition of the paper (*5.2.9 resource type*) and are impressed when they find that it contains lots of information from the same day. The situation ends, after 3 minutes and 30 seconds, when they select an entry in their bookmark file.

The situation is a good example of the complex relationships between factors that influence how searchers interact with the Web. The searchers wish to compare Norwegian digital newspapers with foreign ones, but when they understand that the resource they wish to examine is too different from its Norwegian counterparts, they decide to leave it out. The page is cognitively, but not situationally relevant (Saracevic, 1996b). Note also that the searchers were aware of this resource before actually accessing it, and the fact that they had been told it was a reputable resource made them very interested in examining it.

> The situation (S6_S13) that takes place in Oktober's web resource (*5.2.9 resource type*) reflects the searcher's confusion (*2.8 attention - → 2.1 task knowledge - → 5.2.4 information space*). She selects the link to Oktober's author index (*5.2.1 action*) and finds a page presenting E. Hoem, whom she believes she has found an entry on in another publisher's resource (*5.2.4 information space - →*). The page found in 6_S5 was a presentation of a book of his, from *Oktober*. She is still convinced that K. Holt is published by Oktober and is frustrated (*2.7 uncertainty/affection*) saying "[I] I can't find him here". She decides to "continue with newspapers" and selects Kvasir from her bookmarks (*5.2.1 action*). The situation lasts 1 minute and 6 seconds.

Prior to this situation the searcher has learnt that the author E. Hoem is published by the publisher Oktober. The searcher, however, has forgotten this and believes that K. Holt, which is another author she is looking for information about, is the one published by Oktober. We thus see that the "erroneous" content of the searcher's information space makes her uncertain.

> Aftenposten is the largest Norwegian broadsheet newspaper, and has an extensive web service. She scans and quickly finds Dahl's entry (*5.2.1 action*) in the paper's literature review page (*5.2.9 resource type*) where the authors are listed alphabetically. Here there are links to the paper's review of his books as well as other articles and interviews with him published in the paper. She is well aware of this structure from previous experience (*5.2.2 accumulated results* → 5.2.4 *information space*) and decides to bookmark (*5.2.6 relevance judgement*) this page without looking at the individual articles (← *5.2.7 relevance level* ← *1.1 WT goal*). It is a very targeted situation, which only lasts for 26 seconds. The searcher returns to the index to Norwegian literature reviews via a bookmark (*5.2.1 action*).

> *Like the first two situations S4_S5 is characterised by a searcher who is well aware of where she may find the relevant information. The resources are characterised by representing large, well-known Norwegian organisations (2.1 task knowledge + → 5.2.4 information space + → 4.3 ST strategy). The resources she uses are quite similarly structured, with an alphabetical index to the authors (5.2.9 resource types). On the technical side no problems occur during transaction. One should also note that the transition goes via a bookmarked index rather than a bookmark that leads the searcher directly to the resource.*

The last example, which also includes an interpretation part (in italics), illustrates how the searcher transfers her knowledge about non-digital resources to her information space and decides that resources she thinks will be of use in their original form may well be available and valuable in digital format. The excerpt also shows an example of the relevance level being topical when the relevance judgement is made; the searcher judges the web page relevant based on the knowledge that it contains something about a specific author but without evaluating whether the content is of use to her. We also see that the work task goal is identified as influencing the relevance level, because the searcher's work task is to create a web resource containing information about a set of authors. The web resource is aimed at high school students who are supposed to write their mandatory essays on fiction authors. Thus the author information collected is not directly of use to the searcher, who rather plays the role of the intermediary.

5.5 Analysing stories

The remaining parts of this chapter will be dedicated to the analysis of our empirical data. We start by examining some quantifiable characteristics of the 9 search sessions and, in Subsection 5.5.1, we take a detailed look at search situations and transitions to identify similarities and differences. In Subsection 5.5.2-5.5.6 we concentrate on the empirical research questions stated in Subsection 5.2.

First (in Table 5.4) we present a short summary of the basic features identified using transcriptions of the video. We have looked at the relationship between numbers of actions, i.e. actions performed on/within a resource, and time spent in the individual search sessions.

In total almost 4500 individual actions were identified in the sessions. Sessions 2-8 were all connected to the same work task and with one exception they were all performed by one searcher. One searcher participated in sessions 2-4, another searcher performed sessions 5 and 6, and a third searcher performed sessions 7 and 8. We can see that for the

individual searcher there is a clear connection between the number of actions and time spent. Rank correlation, independent of the searcher performing the session, is 0.72. The

Session no	Category	Participants	Actions	Time	Action rank	Time rank
1	INFO	2	456	1h 44 min	6	3
2	WEB	2	561	2h 6 min	4	1
3	WEB	1	652	1h 27 min	3	6
4	WEB	1	862	1h 52 min	1	2
5	WEB	1	214	30 min[20]	8	9
6	WEB	1	779	1h 37 min	2	4
7	WEB	1	166	44 min	9	8
8	WEB	1	509	1h 30 min	5	5
9	INFO	2	290	56 min	7	7
Total			4489	12h 24 min		

Table 5.4 Characteristics of the 9 Web search sessions. Actions performed during the sessions

table shows that the time spent per action differs, which may be due to several reasons, for example characteristics of the searchers (reading speed, cognitive style, personality etc.), or possibly characteristics of their search tasks or work tasks.

5.5.1 Situations and transitions

Table 5.4 can be supplemented by the numbers of situations and transitions performed in each of the 9 sessions, which is presented in Table 5.5. Each session generates a search *process*, so that when discussing individual sessions we refer to individual search processes as well. Please note that we have not included the end of the session/process as a transition on its own, due to the fact that some of the sessions ended accidentally and not as the result of an active decision made by the searcher. We start our analysis by focusing on the role of situations and transitions. In this analysis we do not focus on the relationship between WIS processes and other categories; we are interested in the "internal" characteristics of the processes and what constitute the observable differences of situations and transitions.

[20] Session 5 originally lasted 2 hours and 5 minutes, but it was unfortunately over-recorded after only 30 minutes.

Situations and transitions are considered to play significant parts in conceptualising Web information search strategies. Searchers can be said to be in different kinds of modes when interacting with information and meta-information respectively. Although we believe that situations and transitions share many characterising attributes, there may be differences in their attributes' relationships. In addition, there may be unidentified or unspecified attributes that play different roles during situations and transitions.

In the present subsection we examine the characteristics of situations and transitions with respect to their frequency and how much time searchers have invested in them. We also look at the frequencies of situation and transition specific attributes in order to see whether these show any patterns.

Session	Situations	Transitions	Time	Situation rank	Transition rank	Situations per Transition
1	28	38	1h 44 min	5	5	0.74
2	18	23	2h 6 min	6	7	0.78
3	51	59	1h 27 min	2	3	0.86
4	64	75	1h 52 min	1	1	0.85
5	17	19	30 min	7	8	0.89
6	49	65	1h 37 min	3	2	0.75
7	8	11	44 min	9	9	0.73
8	33	52	1h 30 min	4	4	0.63
9	15	27	56 min	8	6	0.56
Sum/mean	283	369	12h 26 min			0.75

Table 5.5 The relationship between situations and transitions. Frequency of situations and transitions

In Table 5.5 we can see that in total 283 situations and 369 transitions, i.e. in all 652 "episodes", were identified. This means that 283 recordings have been made of episodes where the searchers access resources with the hope of finding something that could help them to satisfy their information need. Combining the findings reported in Tables 5.4 and 5.5 we see that each episode on average consists of 7 actions.

We can see that there is a correlation between the numbers of situations and transitions. It is interesting to note that the longest-lasting search process consists of comparatively few situations and transitions. This might have to do with characteristics of the work task and/or the search task(s) that these two searchers are involved in. This process, which takes place in session 2, deals with the same work task as in sessions 3-8. We will examine

characteristics of these sessions more thoroughly later. What can be of particular interest when looking at the characteristics of the session, with respect to the number of situations and transitions, is the search task goal. Sessions 1, 3, 4, 5, 6, and 8 all have as their point of departure to collect information about authors. Of these sessions 1 deals with one author while the ST goals generating the remaining search processes were to collect information on predefined sets of authors. If we look at the situation and transition rankings, we may see that the processes generating the fewest situations and transitions are those not having to do with collection information on authors, the interrupted session (S5) being the significant exception to this rule.

In 5 sessions the mean number of situations per transition lies between 0.75 and 0.89, which is surprisingly high – the mean number of all sessions equals 0.75. This indicates that there is a 75 % chance that the use of a search tool, for example a search engine or a subject index, leads to access to a resource that the searcher believes is useful. This seems to indicate efficient searching, and we therefore need to point out a limitation in the transition concept. Our definition of transitions does not take into consideration the fact that searchers often reformulate their queries before finding results worth looking at. In other words, we treat all interactions within a resource as a single transition unless the searcher changes her search (sub)task goal. In addition searchers often choose to re-use the same query result, meaning that when they backtrack from a resource to the query result they often continue by choosing the next entry. Thus the investments of time and effort in the initial transition performed in a resource are undervalued when using this number as a measure of efficiency of search performance. Nevertheless, it constitutes an interesting measure for comparing WIS processes generated by, for example, different kinds of work tasks and search tasks.

In Table 5.6 we present situation and transition characteristics with respect to the time spent in executing them.

Session	Time spent in situations	Time spent in transitions	% of process time spent on situations	Mean length of situations	Mean length of transitions
1	1h 4 min	40 min	61.5	2 m 17 sec	1m 3 sec
2	1h 20 min	46 min	63.5	4 m 27 sec	2 m
3	49 min	38 min	56.3	58 sec	39 sec
4	1h 2 min	50 min	55.4	58 sec	40 sec
5	19 min	11 min	63.3	1 m 7 sec	35 sec
6	1h 4 min	33 min	66	1 m 18 sec	30 sec
7	22 min	22 min	50	2 m 49 sec	2 min
8	54 min	36 min	60	1 m 38 sec	42 sec
9	25 min	31 min	44.6	1 m 40 sec	1 m 9 sec
Average			57.9	1 m 55 sec	1 m 2 sec

Table 5.6 Characteristics of situations and transitions. Time spent in situations and transitions

In 7 out of the 9 search processes we have observed that the majority of time has been used by the searchers to interact with resources where they hope to find information that could help them solve their task. Very roughly it may seem that situations consume twice as much time as transitions in our 9 sessions. In only one session (S9) have the searchers used more time examining meta-information than "real" information. The latter session was also "inefficient" in the sense that the process generated quite a small number of situations compared to transitions.

S9 is characterised by being performed by two searchers, whose tenacity is quite low when it comes to examining query result lists, and they are performing the session at a very early stage in their work task process. Their search task goal is partly unclear and so is the corresponding search task strategy formulated for this search process. If, however, we look at the time spent per situation things appear a little bit different. When taking this perspective we see that S9 situations are longer than those generated by S3-S6 and S8. It is interesting to note that these particular situations are those that we examined as being similar with respect to both their work task and their search task goals. We follow up on the analysis of the role of work task goals below.

In Table 5.7 (see next page) we summarise the search process attributes identified in all sessions, which may help to clarify some of the differences between situations and transitions.

We can see that situations seem to be more complex than transitions. We have already noted that they take up twice as much time as transitions, which is also reflected in the total number of attributes identified in the sessions. For transitions, however, "action" and "resource type" represent 76.5 % of all attribute occurrences, while they constitute only 54.4 % of the situation attributes. Attributes dealing with relevance of course make a big difference, representing 17 % of the situation attributes, but still there is a larger share of other attributes identified during situations than in transitions (28.6 % and 23.5 % respectively) and this would still have been the case even if the relevance attributes were not taken into consideration. If we leave relevance judgement and relevance level out of the statistical analysis it appears that attributes other than "action" and "resource type" represent 34.5 % of the attributes that have been recorded during situations.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	SUM
Transition attributes										
5.1.1 action	68	46	81	100	37	77	27	81	59	576
5.1.2 acc. results	9	8	9	19	6	8	4	5	12	80
5.1.3 acc. effort	8	4	2	8	3	8	2	2	12	49
5.1.4 info-space	13	17	7	8	6	4	7	4	9	75
5.1.5 time	6	7	6	7	1	10	3	4	7	51
5.1.6 remaining needs	3	1	4	5	0	4	0	3	2	22
5.1.7 resource type	43	28	51	79	25	61	17	51	41	396
5.1.8 tech. problems	1	4	3	5	1	3	1	3	1	22
SUM	151	115	163	231	79	175	61	153	143	1271
Situation attributes										
5.2.1 action	54	38	62	129	50	109	29	96	36	603
5.2.2 acc. results	11	11	21	25	11	30	3	14	3	129
5.2.3 acc. effort	10	12	18	16	6	15	3	14	6	100
5.2.4 info-space	25	13	25	35	19	28	7	27	7	186
5.2.5 time	16	11	21	13	6	17	2	12	1	99
5.2.6 rel. judgement	18	18	52	48	15	27	8	36	12	234
5.2.7 rel. level	3	2	34	15	11	21	3	19	1	109
5.2.8 remaining needs	6	6	6	3	0	4	1	1	0	27
5.2.9 resource type	51	25	63	112	29	92	18	70	33	493
5.2.10 tech problems	2	7	5	6	5	5	0	6	0	36
SUM	196	143	307	402	152	348	74	295	99	2016

Table 5.7 Summary of transition (5.1.1 – 5.1.7) and situation attributes (5.2.1 – 5.2.9) for all search sessions (A complete list with definitions of attributes is found in Subsection 5.3.2)

We believe that this can be used to show that situations are somewhat richer in variety than transitions. The situations vary with respect to the resource types that searchers interact with and, since these additionally are resources in which searchers expect to find

information, they choose to invest more time and effort in examining the resources. In addition it seems plausible that searchers are more imaginative and creative during situations in order to get the most out of the resource. In turn this has an additional effect in the form of investment of time and effort, re-use of accumulated results, activation of information space etc. Further analysis should be performed to investigate this.

Our empirical research questions may reveal whether there are causative effects that are particular for situations and transitions. We have looked at how four different factors directly influence search process factors, making it possible to identify effects on transition and situation attributes individually. Further analysis would be needed in order to study the effects of other attributes.

5.5.2 Empirical research questions

To find answers to our research questions we could choose to identify each occurrence where one of the relevant attributes (work task goal, work task stage, search task strategy, work task knowledge, and search knowledge) influences one or more attributes of the search process. We may then see whether specific relationships are of particular interest and what implication this may have.

This is, however, not the only way to do it, at least for the analysis of work task influence. An alternative option would be to categorise the sessions according to the goal of the work task or the stage of the work task performance the searchers are in. Selected attributes may then be focused on to see how they relate to one another in sessions differing according to, for example, work task stage. The latter alternative is used to analyse the influence of work task stage, while we have performed a "micro-level" analysis for the other empirical research questions.

Please note that the research questions have been raised to study the method schema's applicability to the problem type, not to get precise answers concerning the relationships on a general level.

5.5.3 Influence of work task goals on WIS processes

On a general level the goals of all the students we have observed searching the Web have been to complete their theses. We are interested in finding out in what instances the work task goals directly influence the web information search (WIS) process. In most cases we

believe the WT goal will only *indirectly* influence the searching, via its core role during the formulation of the search task goal(s). In Table 5.8 we present the relationships found in our stories, where the WT goal has been identified as influencing the search process.

1.1 work task goal →	S1	S2	S3	S4	S5	S6	S7	S8	S9	Sum
Transition attributes										
5.1.6 remaining needs									1	1
Situation attributes										
5.2.6 relevance judgement		1					1		1	3
5.2.7 relevance level	1	1	18	23	5	20		11	1	80

Table 5.8 How the work task goal affects the WIS process

Clearly the goal of the work task directly influences the relevance level used by our searchers when they evaluate resources. The sessions where this influence is most apparent are all generated by the same work task, which was to develop a web resource on a predefined set of authors targeted at high school students writing their theses. These particular sessions (S3, S4, S5, S6, and S8) also have similar search tasks, i.e. to collect information on authors to include in the resource. The work task thus makes the searchers evaluate the information found according to other criteria than if they had needed the information for literary analysis or biographic storytelling. The criteria used by the searchers reflect that their relevance level is *topical* (Saracevic, 1996b). By this we mean that the searchers are only interested in finding out whether the retrieved web page contains "something", such as an interview or a book review, about the author on their list. They do not analyse the contents of the page any further, but leave this to the "end users". It is our claim that the searchers take the intermediary role when evaluating documents representing authors in these 5 sessions.

One reason why we have not been able to identify the influence of work task goals on relevance level in the other sessions may be because we take for granted that the relevance level is what Saracevic (1996b) calls *situational* relevance (see also the rules for classifying relevance level in Appendix 3). By this we mean that searchers judge potential information by its usefulness in reducing uncertainty. This is not the case in these five sessions, where the relevance level was "topical", meaning that the searchers examined whether the resources were about the author without relating them to any specific problem. It is easier to spot what was not predicted in the data, and in this case it was predicted that situational relevance would be the searchers' default relevance level.

On the other hand we hypothesise that the possible users of the resource, i.e. the high school students, will evaluate whether the potential information in the Web site is *cognitively*[21] (Saracevic, 1996b) or situationally relevant.

It could be argued that in the particular cases referred to above – searchers taking the intermediary role – it would be sufficient to use the *search* task goal to perform the relevance judgement. While in cases where the searchers search information in order to satisfy their situational needs they evoke the work task goal in order to decide whether the retrieved pages can be of use to them. On the other hand, Fidel and her colleagues (1999) have found that high school students use criteria such as the graphics used on a page or the amount of information to evaluate the usefulness of Web pages, suggesting that there is clearly a need for more research on the issue. This discussion generates the following hypotheses, which would be interesting to test:

1. Search task goal can be used to perform relevance judgements on the topical level;
2. Work task goal can be used to perform relevance judgements on the cognitive level;
3. Work task goal can be used to perform relevance judgements on the situational level.

We leave these hypotheses for future research in order to concentrate on other characteristics of the work task goal – relevance level-relationship.

There are examples of how searchers change parts of their work task goal as the result of lack of relevant sources, as for example the last paragraph in the excerpt below. Here the searcher decides not to include a certain page on a certain author, while similar pages have been used to represent other authors. Thus the accumulated results will influence the searcher's knowledge of the task, which may make her adjust her work task goal and in turn use new relevance criteria. This example of how work tasks are dynamic, influencing the relevance judgements as well as the direction of the WIS process as a result of the retrieved documents, supports claims often made by theorists on information behaviour (e.g. Ingwersen, 1996; Vakkari, 2001). Our method schema can thus be used to make methods that are able identify such instances in WIS processes provided that coding is done in sufficient detail.

Indirectly the work task goal also influences the time (and effort) invested in examining resources, when judging their relevance during the process. In addition the actions

[21] Cognitive relevance, or pertinence, is used by Saracevic to express "the relation between the state of knowledge and cognitive information need of a user, and texts retrieved". At this level there is, however, no attempt at evaluating the relationship between the searcher's work task and the retrieved documents.

performed in order to inspect the page are also influenced. The searchers chose to invest the time in examining as many resources as possible rather than thoroughly looking into the individual resource. This explains the findings from above, where we saw that the search processes that dealt with collecting author information for the student resource generated the shortest situations.

We show an excerpt of our data below.

> During transition S5_T10 the searcher explains that at this stage she does not thoroughly read the texts she finds, but tries to get an overview (*5.1.4 information space → 5.1.1 action*). She finds and selects a link in Kvasir's query results (*5.1.1 action – 5.1.7 resource type*) to an interview with Gaarder. (17 secs.)
> The interview (*5.2.9 resource type*) scanned (*5.2.1 action*) and bookmarked (*5.2.6 relevance judgement*) in situation S5_S9 is included in the final resource and the searcher comments that it looks interesting. She spends 25 seconds on the situation (*5.2.5 time* affected by the searcher's decision to aim only for an overview ← *5.2.7 relevance level*, which in turn is affected by the ←*1.1 WT goal*) before backtracking to the query results.
> She comments that when she adds many bookmarks (*5.2.2 accumulated results*) she has problems in deciding whether a page has already been bookmarked – i.e. a kind of cognitive overload (*5.2.2 accumulated results → 5.2.4 information space - → 2.8 attention*) influences the session. She explains that one strategy she has used previously to deal with this is to print (*4.3 ST strategy → 5.2.1 action*) all bookmarked pages to look for duplicates. This she finds easier than working only at the bookmark entry level.
> Transition S5_T11 (24 secs.) is similar to the previous transitions (*5.1.2 accumulated results*) dealing with the Kvasir query results (S5_T8, S5_T10) as well as to the forthcoming (S5_T12 and S5_T13); the searcher scans the query results page (*5.1.1 action – 5.1.7 resource type*) and selects (*5.1.1 action*) a link.
> In situation S5_S10 the resource contains a presentation of a novel (*5.2.9 resource type*) by Gaarder called "I et speil i en gåte". Gaarder's publisher Aschehoug has made this page, which is bookmarked, but not included in the final resource (*5.2.6 relevance judgement*). The situation lasts only 13 seconds (*5.2.5 time*) when the searcher backtracks to the query results list (← *5.2.7 relevance level ← 1.1 WT goal*).
> *It is clear from examining the final resource that different criteria have been used when the students have chosen which pages they think are relevant for the individual authors* (*5.2.2 accumulated results → 2.1 WT knowledge → 1.1 WT goal → 5.2.7 relevance level → 5.2.6 relevance judgement). One possible reason for this is the amount of information available on the individual authors. The case of Fangen and Gaarder is a good example. Gaarder is a contemporary author who has received much attention internationally and a phrase query on his name in FAST December 2000 receives 5288 hits. A similar query on Ronald Fangen returns 111 hits, which is not bad considering he has been dead for fifty years and was not considered to be that "great" during his period of writing. On the other hand Fangen is not included at all in the final resource, but many other authors are represented with their publishers' presentations. This indicates that the amount of available information on a topic directly influences the relevance criteria used.*

The example shows two situations (S5_S9 and S5_S10) where the searcher evaluates pages for inclusion in the resource and two transitions (S5_10 and S5_T11) during which the searcher inspect results from a query in the Norwegian search tool Kvasir. The final paragraph of the excerpt is an example of how we have interpreted the context of the process; it has been useful to add such interpretations separately if they cannot be included naturally in the descriptive part of the story.

5.5.4 Influence of search task strategies on WIS processes

In our story analysis we have been interested in investigating relationships between attributes that we believe are particularly open to the dynamics of search processes. The search task strategy is one such attribute, since the strategy may change as the results of what the searcher finds or does not find, and it may depend on time spent and effort invested. To investigate the role of search task strategies we have therefore not only looked at how the strategies influence other factors but also how the strategies themselves are influenced during the search process.

We start by examining the influence of ST strategies on other attributes during the WIS process. Table 5.9 summarises our findings, where we first present search process attributes that were affected and then attributes of other categories. The relationships marked "+" mean that the ST strategy had a positive effect on that attribute, thus "5.2.5 time +" means that the ST strategy can be identified as being time-saving. On the other hand "5.2.5 –" implies use of a ST strategy that made the process/situation last longer than it needed to.

4.3 ST strategies →	S1	S2	S3	S4	S5	S6	S7	S8	S9	Sum
Transition attributes										
5.1.1 action	6	2	3	6	3		1	1	2	24
Situation attributes										
5.2.1 action	2	2	2	6	2	5		3	1	23
5.2.5 time +						1				1
5.2.5 time -						1		2		3
Searcher attributes										
2.7 uncertainty		1			1					2

Table 5.9 How search task strategies affect the WIS process

Not surprisingly the result of the analysis shows that search task strategies directly affect the actions taken during the search process, both during situations and transitions. In

most cases actions are dependent on the resource type within which they are performed. Therefore it is reasonable to say that the search task strategy also affects what resource (type) to use. This is not reflected in the table due to the way we have treated the "action-resource type"-relationship (please refer to Subsection 5.4.6). It is also clear that well-formulated search task strategies save the time (and effort) spent during the process, and vice versa, i.e. bad ST strategies cause inefficient use of time. The following situation illustrates the latter:

> In S8_S2 the searcher scans the Gyldendal author biography index (*5.2.1 action – 5.2.9 resource type*) for approximately 1 minute and 20 seconds without finding any of the needed authors. She comments that she should have done research beforehand to learn each author's publisher (*2.1 WT knowledge - → 4.3 ST strategy - → 5.2.5 time & 5.2.3 effort*). In the end she selects a link to an index of external resources on authors published by Gyldendal. This index is also scanned (*5.2.1 action – 5.2.9 resource type*) for a considerable amount of time (approx. 45 seconds) before she selects a link (*5.2.1 action*) to a presentation of Agnar Mykle made by Aftenposten Alex. The selection of the link is categorised as a transition to a new knowledge source (S8_T3 – 4 secs.) and ends situation S8_S2 after 2 minutes and 12 seconds.

The situation exemplifies a searcher who has developed an ST strategy without having the necessary WT knowledge, i.e. knowledge about an author's publisher, which makes her spend time and effort trying to identify the correct publisher using different publishers' web resources.

The next example is collected from the same session, and shows two examples of how the ST strategy influences the actions performed.

> In situation S8_S8 the searcher looks at the top part of the index (the newspaper Aftenposten's index to pages presenting authors) for more than 20 seconds before she decides to select the link (*5.2.1 action – 5.2.9 resource type*) that leads her to the part of the index that contains authors with names that start with an "H" (*5.2.9 resource type*). From this position she scans (*5.2.1 action*) down until she finds and selects the entry (*5.2.1 action – 5.2.9 resource type*) on Roy Jacobsen. She examines (*5.2.1 action*) the Jacobsen page (*5.2.9 resource type*) and selects a link (*5.2.1 action*) to an interview (*5.2.9 resource type*) before she backtracks (*5.2.1 action*) and bookmarks the page (*5.2.6 relevance judgement*). She next follows exactly the same procedure (*5.2.1 action – 5.2.9 resource type…5.2.1 action …5.2.6 relevance judgement ← 5.2.7 relevance level ← 1.1 WT goal*) for the other authors she is looking for that are represented in the index (Loe, Michelet, Mykle, and Nedreaas), stating "I just take the list down here" (*4.3 ST strategy → 5.2.1 action – 5.2.9 resource type*). The situation lasts 3 minutes and 42 seconds and must be categorised as very efficient and probably satisfactory for the searcher (*5.2.2 accumulated results + → 2.5 motivation ← + 5.2.3 accumulated effort*). The situation ends when she selects the Kvasir bookmark entry (*5.2.1 action*).

The searcher next enters the query "dagbladet" in Kvasir (*5.1.1 action – 5.1.7 resource type*), which is the name of a large Norwegian tabloid newspaper with a quite good cultural reputation. She scans the query results (*5.1.1 action – 5.1.7 resource type*) although she quickly finds the link she wishes to follow. The following of this link (*5.1.1 action*) starts situation S8_S9 and ends transition S8_T9. (34 secs.)

Situation S8_S9 starts in Dagbladet's literature page, which, the searcher comments, "consists of today's [literary] news" and "is not similar to Aftenposten [Alex]" as she had thought (*5.2.4 information space ←→ 5.2.9 resource type*). She selects a link (*5.2.1 action*) that leads to an index to book reviews (*5.2.9 resource type*), but does not bother to spend much time in it as it is organised chronologically rather than by author (*5.2.4 information space - → 2.5 motivation - → 5.2.3 effort*). She finds a link named "søk" (search) and follows this (*5.2.1 action*) to the query page where she enters the query "jacobsen, roy"(*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action – 5.2.9 resource type*). The query's bibliographical form does not match Dagbladet's index routines and the query is treated as a Boolean OR query, thus retrieving many non-relevant entries in the query result list. The searcher follows the first entry (*5.2.1 action*) in the list, which leads her to the visit card (*5.2.9 resource type*) of one of Dagbladet's employees, whose first name is "Roy". She also looks at the second entry (*5.2.1 action*), which is a short note (*5.2.9 resource type*) containing a statement of someone called "Roy". The searcher decides to enter a new query, as she is unable to find any relevant entries. The next query entered is "loe, erlend" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action – 5.2.9 resource type*), and as this name contains terms that are much less frequent and most probably because the author at the time of the session was very much written about, the query generates three pages containing information about the author, all of which were accessed (*5.2.1 action*) and of these one was bookmarked (*5.2.6 relevance judgement*). The fourth link followed turns out to be a review of a book by another author named "Erlend". The searcher comments that the result list format is "bad, it says nothing about what the [links contain]" (*5.2.9 resource type ←→ 5.2.4 information space*). The searcher followed links that were categorised as "kultur" (culture) or anmeldelser (reviews) (*2.1 WT knowledge + → 4.3 ST strategy → 5.2.1 action*), but the links themselves only consists of the files' alphanumeric name! The searcher continues to use this *strategy*, i.e. looking for articles in the results list categorised as "culture" or reviews in the continuation of the situation. She searches for "michelet, john" (his real first name is "jon" – *2.1 WT knowledge*), "michelet, j", "nedraas, torborg" (misspelling of "nedr**e**aas" – *2.1 WT knowledge*), and magerøy, ragnhild", (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action – 5.2.9 resource type*) which all in all results in accessing 10 different pages, of which 5 are to a certain degree topically relevant and 1 is bookmarked (a page on a theatre play based on a book by Nedreaas, correctly spelt!) (*1.1 WT goal → 5.2.7 relevance level → 5.2.6 relevance judgement*). The situation lasts 10 minutes and 18 seconds. The new transition starts when the searcher selects Kvasir from her bookmarks.

In situation 8 she decides that it is rational to collect information on all authors at one time. The format of the resource facilitates such a strategy and, as we have discussed above, the work task of this searcher is one of an intermediary's; she collects and presents information that is to be used as a reservoir for future searchers with other work tasks.

For this reason the level of relevance is topical and the searcher may choose a ST strategy that is efficient for dealing with information satisfying that level.

In the second situation (situation 9) we see that the searcher has to adapt her search task strategy to deal with a resource unfamiliar to her. Having analysed the format of the query result list, she understands that it is a good idea to focus on those entries dealing with culture or reviews, since apparently her query retrieves lots of documents that have nothing to do with the person she is looking for. If her search knowledge had been better, and not as focused on the "bibliographic format" as it is, she might have been able to formulate a Boolean AND-query. In fact a correctly formulated AND-query would also have improved the format of the entries in the query result list; the entries would then contained extracts from the indexed documents where the query terms would have been emphasised.

The latter situation also exemplifies how the ST strategy is influenced by other attributes. The searcher's work task knowledge causes her to be aware of the entries that most probably contain information about authors.

We will next look at the attributes that affect the ST strategies. In Table 5.10 we present the relationships that we have found where ST strategies are influenced by attributes of our framework's five categories. We list the attributes starting with the search process and then the other categories in "numerical" order. This order is chosen in order to emphasise the direct effect of the attributes of the WIS process. Some relationships are marked with plusses and/or minuses. We use "+" to indicate that the effect is positive, for example that the searcher's WT knowledge makes her formulate a suitable ST strategy and vice versa: the searcher's lack of WT knowledge results in a weak ST strategy. The "2.1 WT knowledge -+ → 4.3 ST strategy"-relationships means that the lack of WT knowledge actually has a positive effect on the ST strategy!

→ **4.3 ST strategies**	S1	S2	S3	S4	S5	S6	S7	S8	S9	Sum
Transition attributes										
5.1.2 accumulated results	1		2	1	1	2		1	1	9
5.1.3 accumulated effort					1	1				2
5.1.4 information space	2	3	2	1	1	1	1	1	2	14
5.1.6 remaining needs				3		1				4
Situation attributes										
5.2.2 accumulated results		2		3	2	1	1	1	1	11
5.2.3 accumulated effort			1		1					2
5.2.4 information space			1	1		3	1	3		9
5.2.5 time	1							1		2
5.2.8 remaining needs	2	2	1	1		1				7
Work task attributes										
1.1 WT goal			1			3	3		1	8
1.3 WT stage									1	1
Searcher attributes										
2.1 WT knowledge +	6			4		1		3	1	15
2.1 WT knowledge -	2	2		1	3	2	1	2	1	14
2.1 WT knowledge -+									2	2
2.2 search knowledge +				1	1			2		4
2.3 search system knowledge +								1		1
2.5 motivation +							1			1
2.5 motivation -		1		2		1			1	5
2.7 uncertainty -								1		1
2.8 attention +						1				1
2.8 attention -	1					2				3
Environment attributes										
3.2 actors	1					1				2
Search task attributes										
4.1 ST goal	1	3	1			1			4	10
4.2 ST complexity	7		2	1	1	1			1	13

Table 5.10 Category attributes that affect search task strategies during WIS processes

We see that the two most important characteristics of the search process when it comes to influencing search task strategies are the information space and the accumulated results. This should not come as a surprise. In addition WT knowledge is an important influence. We shall return to this relationship in the next subsection of this chapter.

The information space is the searcher's model of the Web, which contains correct and erroneous "knowledge" about its content, structure, and interaction techniques. This model may, for example, have a representation of a prototype publisher resource, in other words it may model how any publisher's web resource is structured with respect to its navigation facilities, its information architecture, the kind of pages it provides (index to the authors, index to different genres etc.). The information space may provide the searcher with ideas about how to attack her problem, for example "look for author A in

publisher P's resource, if this does not work proceed to publisher R, etc." or "look for authors A-Z in publisher P's resource, if some of these authors are not found proceed to look for the remaining in publisher R's resource" etc. Below we give two examples of how the information space influences the ST strategies in the data we have analysed.

If we consider the difference between situations and transitions it may be interesting to note that searchers were identified as more often using their information spaces to influence their ST strategies during transitions (code 5.1.4) than in situations (code 5.2.4). This means that searchers use their web model when they plan how to use the search tools in a satisfactory way. This, in turn, indicates that ST strategies are not clearly formulated prior to the sessions, but are the result of dynamic processes.

In the example below we step into the search process at quite a late stage where the two searchers return to a resource that they have previously used to find graphics for their project. Now they wish to use it in another way:

> Icon Bazaar is yet again the resource used, but this time for transitional purposes (S2_T22). They believe it might be possible to find some usable resources by examining Icon Bazaar's own link collection to "other collections" (*5.1.7 resource type → 5.1.4 information space → 4.3 ST strategy*). Unfortunately the server does not respond when they activate the link, which they do twice and they spend 2 minutes and 6 seconds waiting for the page to turn up before giving up (*5.1.8 technical problems → 5.1.5 time - → 2.3. motivation*). The searchers therefore select another bookmark entry (starting S2_T23).

It is clear that these searchers have a model of how the resource is built up and that they wish to perform certain actions within the resource in order to find links to potentially useful resources. Due to technical problems their strategy fails, but it nevertheless exemplifies a relationship between the information space and a search task strategy. The next example is collected from a situation.

> The page used as starting point in situation S6_S18 is VG's index to cultural resources (*5.2.9 resource type*) in which she selects the link (*5.2.1 action*) to "books". Data transfer is very slow (*5.2.5 time*) and the searcher interrupts the data flow twice before at the third try (*5.2.10 technical problems - → 2.5 motivation ← + 2.6 tenacity*) she is able to access the book page. She has to select one more link to get to the actual book reviews, and it takes approximately 2 minutes and 20 seconds in transfer *time* to get to this page from the start page. When she enters Vg's book review index she comments that it is sorted chronologically (*5.2.4 information space →*) and indicates that she looks for a search feature (*4.3 ST strategy*), which she finds a link to. In the search page the searcher combines the predefined categories "debatt" (debate) and "VG direkte" (*5.2.1 action*), for which she states as a reason that she thinks Anne Holt has had a debate on crime literature in this newspaper (*← + 2.1 WT knowledge*). For some reason the

> query request gets no response, which makes the searcher go to the help page (*5.2.1 resource type*). She explains that she visited the help page the previous day as well without getting any relevant advice (*5.2.2 accumulated results → 5.2.4 information space*). She spends quite a lot of *5.2.5 time* and *5.2.3 effort* scanning the help page (*5.2.1 action – 5.2.9 resource type*), but nothing relevant turns up. She explains that she misses an overview of book reviews (*5.2.4 information space → 2.7 uncertainty/affection*), and she is not even sure that VG has that many book reviews (*2.1 WT knowledge - → 5.2.4 information space*), as she is unable to find them. She then tries to re-enter the query from above (*5.2.1 action*), but there is no response. In the end she selects a link to the paper's consumer related section (*5.2.1 action – 5.2.9 resource type*), but is unable to find anything of relevance in this page nor in the archive on previous consumer related articles (*5.2.9 resource type*). She ends the situation by first backtracking (*5.2.1 action*) several times before selecting a bookmark entry (*5.2.1 action*) to the literature resource index. She says that she gives up using this resource (*5.2.4 information space → 2.5 motivation*) and will ask her fellow students (*3.2 actors*) if they have been able to use the paper and can give her any good ideas. The situation lasts 11 minutes and 13 seconds.

We see that the searcher reacts to the structure of the resource by searching for a familiar feature of similar resources, i.e. the local search engine. "VG" is Norway's largest tabloid newspaper and she obviously has a clear idea about how such resources are structured. Prior to this situation she has spent much time and effort querying for authors in Dagbladet, which is the second largest tabloid newspaper. It is therefore reasonable that she would choose a similar strategy in this resource. She does, however, experience problems in using the database and is unable to learn how to use the resource to find author information. In the end she gives up and says that she will discuss how to use this resource with her fellow students.

The accumulated results influence the ST strategies in a similar way as the information space; if the searcher has found the desired information in resource A, she may choose to perform the necessary adjustment in her strategy to continue with another search sub-task. In other words one can say that successful results motivate the searcher to continue using similar strategies and likewise that lack of success may make her change the strategy. We show one example of the accumulated results-ST strategy relationship below.

> In S9_T12 yet another index to interface research (*5.1.7 resource type*), called the UI index, is explored (*5.1.1 action*). The searchers obviously try to learn more about the topic (*2.1 WT knowledge -+ → 4.3 ST strategy*) and also check out a page presenting the originators of the resource. After finding a link collection that points to universities that perform research on HCI, they decide to revise their *4.3 ST strategy* in order to solve their adjusted ← *4.1 ST goal*, probably because the information presented in this resource makes them realise that their needs are at a "lower level" (*5.1.3 accumulated results → 2.1 WT knowledge + → 1.1 WT goal*) than what is presented here (51 seconds). They end the transition by selecting a bookmark (*5.1.1 action*) to Nielsen's Web Alert box, used in S9_S1, to see what topic it covered (or what words

> it used to represent its topic - i.e. usability) (*5.1.2 accumulated results → 4.3 ST strategy*). (51 secs.)

In addition to showing how the searchers plan to use the accumulated results in order to formulate their ST strategy, the excerpt shows that they need to learn more about their topic and perhaps to reformulate their WT goal. The example is a prototype example of how searchers can use what they have experienced during a search process to make their ST strategy clearer, thus supporting the observation by many that information needs are dynamic (Ingwersen, 1996; Borlund, 2000; Vakkari & Hakala, 2000; Vakkari, 2001).

Tables 5.8 and 5.9 show that we have been able to identify many relationships that show how ST strategies influence the search process and how the same strategies are being influenced by what happens during the WIS process. We have shown examples of such relationships and believe that the findings are convincing as part of the justification of our method. We believe that the role of the ST strategy should be further investigated, since this may increase our knowledge about how to become more efficient as Web searchers and may have implications for the design of client software.

5.5.5 Influence of work task knowledge and search knowledge on WIS processes

There have been several studies focusing on search knowledge and domain knowledge influence on information searching Marchionini, Lin & Dwiggins, 1990; Hölscher & Strube, 2000). For this reason it seemed natural to see what relationships could be identified studying our WIS processes.

In Table 5.11 (next page) we summarise the identified relationships where the searchers work task knowledge have been found to influence on other attributes during the WIS processes.

It should come as no surprise that the searcher's WT knowledge affects her model of the World Wide Web (the information space). The searcher constructs her Web model by predicting the kinds of information available on the Web, necessarily believing that it contains information related to her task, or else she would not have bothered to use the Web in the first place. The searcher's WT knowledge is an important point of departure for the development of her search task strategy (see below), and the part of the information space that is put into use is, as the relationship analysis shows, related to the WT knowledge. This suggests that an important part of the information space is that

which the searcher predicts contains the knowledge already possessed by her. Finding the parts of the Web reflecting the information space can then be used as a valuable point of departure for expanding the relevant part of the information space, in other words finding information relevant for her task.

2.1 WT knowledge →	S1	S2	S3	S4	S5	S6	S7	S8	S9	Sum
Transition attributes										
5.1.1 action +	3		2	2		2		1	1	11
5.1.1 action -	1	3	3			1		1		9
5.1.4 information space +	3			1	1					5
5.1.4 information space -							1			1
Situation attributes										
5.2.1 action +	3					2	1			6
5.2.1 action -						2		2		4
5.2.4 information space +	2	1		1	1	2		1	1	9
5.2.4 information space -						2	1			3
5.2.6 relevance judgement +								1	1	2
5.2.6 relevance judgement -	1							1		2
Work task attributes										
1.1 WT goal +					1				2	3
Searcher attributes										
2.2 search knowledge -					1					1
2.5 motivation +	1		1	2		1			1	6
2.7 uncertainty/affection -									2	2
Search task attributes										
4.3 ST strategy +	6			4		1		3	1	15
4.3 ST strategy -	2	2		1	3	2	1	2	1	14
4.3 ST strategy –+									2	2

Table 5.11 How WT knowledge affects other categories/attributes during the WIS process

The formulation of the search task strategy can be seen as a parallel process. The searcher uses, among other things, her WT knowledge and Web model to construct possible ways to find the relevant parts of her information space – the procedure for getting there is the search task strategy.

We see that WT knowledge can have both a positive and a negative effect on the information space as well as on the search task strategy. The negative effect is caused by the searcher having too little WT knowledge to be able to clearly formulate her search task strategy. If the amount of WT knowledge is too small, it may also influence her ability to create a functional model of the Web's coverage and structuring of the kind of information she is looking for. The two cases of "-+" influence on ST strategies in S9 indicate that we have recorded that the searchers become aware of a specific lack of WT

knowledge which in turn makes them formulate a specific strategy for increasing that particular knowledge.

If we compare the roles of situations and transitions for the influence of task knowledge, we may see that half of the process-attributes affected are characteristics of search situations. To learn more about the roles of transitions and situations we examined the stories in order to find out when relationships between WT knowledge and other attributes are identified. This revealed that WT knowledge was identified as influencing other attributes 45 times during situations and 50 times during transitions. Please note that this does not mean that WT knowledge plays a role in 45 *different* situations since there are instances where WT knowledge influences more than one other attribute during the same situation/transition. There are however large differences between the individual search processes, for example 6 out of 7 recorded relationships in S5 took place during situations while in S3 none have been recorded.

In addition to "information space" and "ST strategy", work task knowledge also seems to heavily influence the actions performed during the WIS process. Such relationships have been recorded when the searcher formulates queries signifying WT knowledge or when she has decided to follow links where the link tag apparently grabs her attention because it matches her knowledge. We do not show examples of such relationships here, but choose to show two examples of how WT knowledge influences the information space and ST strategy in two different sessions.

Prior to the transition reproduced below the searchers have unsuccessfully throughout a number of transitions tried to find a Web resource presenting "Bokklubbens barn".

> They (in S1_T19) are apparently reluctant (*2.5 motivation*) to give up their quest for the book club (← *5.1.6 remaining needs*), being most probably influenced by a feeling that such an important organisation must be available on the Web (*5.1.4 information space* → *2.5 motivation*). They next enter the simple query "bokklubb" (*5.1.1 action – 5.1.7 resource type*) in Kvasir. This returns many topically relevant hits, since several book clubs are represented on the Web among those being "Barnas egen bokklubb" (the children's own book club). Their primary goal is still to find "Bokklubbens barn" and they are quite certain that this refers to another book club for children (*2.1 WT knowledge + → 5.1.4 information space +- →*). However they follow the link (*5.1.1 action*) to this other book club for children in case it might contain something of interest to them (probably not a part of their initial *4.3 ST strategy ← 5.1.2 accumulated results*), thus starting a new situation.

We see here that WT knowledge influences the searchers' information space with respect to their expectations of content on the Web. At this point of time the resource they

looked for had not yet published anything on the Web. Note also how their motivation is influenced by their information space.

> They select Internet search from Netscape's embedded directory (*5.1.1 action – 5.1.7 resource type*) (S2_T10), which leads them to Netscape's net search page where they select Lycos. They enter the (simple) query "icon" (*5.1.1 action*), exemplifying little ← - *2.1 WT knowledge*, but when they get the results they erroneously (*2.8 attention* - →) select a link embedded in a commercial banner and get rather confused (*5.1.1 action* - → *5.1.4 information space* - → *2.7 uncertainty/affection*) when they end up in a virtual casino. They quickly understand what went wrong and backtrack to the result list (*5.1.1 action*). Next they scan the result list (*5.1.1 action – 5.1.7 resource type*) and find a link to a Lycos subcategory on "computers". They now try to narrow down their query to be related only to computers, but they are not certain how this should be specified. Their discussion clearly reveals a lack of knowledge about how to attack their problem by using ambiguous terms, e.g. "maybe this [points at link to the Lycos-category **pictures-icon**] is the one that is the most [relevant]" (*5.1.1 action ← 4.3 ST strategy ← - 2.1 WT knowledge*). They scan the computer category (*5.1.1 action – 5.1.7 resource type*) for some time (approximately 1 minute), clearly uncertain about how to explore the resource (*2.7 uncertainty/affection ← - 4.3 ST strategy*) before deciding to re-enter the query "icon" (*5.1.1 action*). The query returns a link to "Anthony's icon library" – last visited (accumulated results) and found *non-relevant* in S2_S8! They do however follow this link and start situation S2_S9. This transition lasts approximately 5 minutes.

Not only does this example show how searchers with fairly little WT knowledge have problems formulating precise queries due to the use of ambiguous query terms, but it also exemplifies how this makes it difficult for them to formulate a clear search task strategy.

2.2 search knowledge	S1	S2	S3	S4	S5	S6	S7	S8	S9	Sum
Transition attributes										
5.1.1 action +	2		4	3					2	11
5.1.1 action +-	2		5		3	14		19		43
5.1.1 action -	1	1		2	1					5
5.1.1 action -+	2		3	7						12
5.1.4 information space +	3									3
Situation attributes										
5.2.1 action +	2			2		4		7		15
5.2.1 action +-	1		1	3	4	12				21
5.2.1 action -	1	1			1					3
5.2.1 action -+				2						2
5.2.4 information space -					1					1
5.2.5 time -						1				1
5.2.6 relevance judgement -			2							2
Searcher attributes										
2.7 uncertainty/affection -					1					1
Search task attributes										
4.3 ST strategy +				1	1			2		4

Table 5.12 How search knowledge affects other categories/attributes during the WIS process

We now go on to look at how searchers' *search* knowledge exercises an influence during WIS processes, a summary of which is found in Table 5.12.

We can see from the table that search knowledge directly influences the actions performed during the WIS processes we have recorded. It is interesting to notice that very many cases have been identified where the searchers have not been able to use their repertoire of search knowledge to perform actions that would help them with their task, but rather the opposite. This is caused by a special characteristic of the searchers we have observed; they are trained in searching bibliographical databases and use the same syntax independent of the kind of database they use. In S1, S3-S6, and S8 the searchers are looking for information about specific authors and, influenced by their LIS education, they formulate their author queries in inverted form even though they use search engines like Alta Vista and subject indices like Yahoo. For this reason there are many "5.1.1 action +-" relationships recorded meaning that their search knowledge impairs their performance. The same syntax is also used when they enter queries in other kinds of web resources, for example in newspaper archives.

When comparing these findings with what we know about the frequency of situation and transition attributes (see Table 5.7), we see that search knowledge influences the actions performed relatively more often during transitions. This can be explained by the fact that searchers' actions during transitions to a large extent have to do with searching in the form of query formulation. Actions performed during situations are more varied; searchers are then more exploratory in the manner of following links and scanning pages.

The result of their "erroneous" syntax use is that their queries are interpreted as Boolean OR-queries. The queries therefore, in some instances, generate very large numbers of hits. If the queries had been formulated using Boolean AND, the searchers would have saved time and effort in handling smaller sets of query results. We show a few examples from the data below.

Example from Session 3:

> From Netscape's search page the searcher selects a link to the search engine (*5.1.1 action 5.1.7 resource type*) Alta Vista, starting S3_T2. What follows is a series of transitions and situations started by the searcher entering the query "Ambjørnsen, Ingvar" in Alta Vista (*5.1.1 action – 5.1.7 resource type*) . She calls the query a "general query" (← *4.3 ST strategy*), but her intention is to find information about a specific author (*4.1 ST goal*). It is probably her background in library science that makes her formulate the query in its bibliographic form (*2.4 education +- → 2.2 search knowledge +- → 5.1.1 action – 5.1.7 resource type*). It would have

been more efficient to formulate it as a phrase, which would also strongly decrease the number of hits since her original query also returns individual hits on "ingvar" and "ambjørnsen" (possible *5.1.1 action*) (1min. 11.secs.).

Example from Session 6:

The searcher (in S6_S1) enters Gyldendal (*5.2.9 resource type*) with the intention "to check out if any of the authors might be found there" (*4.3 ST strategy → 5.2.1 action*). Having scanned the resource's main page she enters the query "hagerup klaus" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action*), which is the name of one of the needed authors in inverted form. The query generates no hits, which indicates that Gyldendal does not publish him. When the searcher returns to the main page the query form has for no apparent reason disappeared (*5.2.10 technical problems → 5.2.5 time*). The searcher manages to reload the page with the query form with the help of the page's redirect feature (*2.3 search system knowledge → 5.2.1 action*). Queries on "hoem edvard" and "henriksen vera" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action*) return no hits. This makes her select a link (5.2.1 *action*) to the index to titles published in 1997. She scans the index (*5.2.1 action – 5.2.9 resource type*) without finding any relevant entries and decides to enter a page that is at the end of a link (*5.2.1 action*) tagged "Katalog". In this page she performs another query (*5.2.1 action – 5.2.9 resource type*). She asks herself whether she should query for title or authors (*4.3 ST strategy*), since "it is a book catalogue", she decides to query on author "hagerup klaus" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action*), but yet again receives no results. The same is the result of new queries on "Hoem" and "Henriksen". She explains that she was not certain what part of the resource she searched in when she used the query formula on the main page, "it might be […] on those who work in Gyldendal" (*5.2.9 resource type ← - 5.2.4 information space - → 2.7 uncertainty/affection → 5.2.1 action → 5.2.5 time + 5.2.3 effort*). She finally suggests that this publisher may not publish the authors she searched for, and that may be the reason for her lack of success. She enters one more query (*5.2.1 action*) before ending the situation, this time on "holth anne" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action – 5.2.9 resource type*), but she is uncertain of how to spell the name: "with [or without] th?" (← - *2.1 WT knowledge*) The situation lasts 4 minutes and 22 seconds and ends when the searcher plans to return to Kvasir via her bookmarks, but finds and selects a bookmark entry (*5.1.1 action*) to a page published by Aschehoug.

Example from Session 8:

When she returns to Kvasir for transition S8_T13 (59 secs.) she selects a new *strategy*; she enters the name of an author in the query field. She says she wants to see what she can find only by using Kvasir. Her query is formulated as "Hussain, khalid" (*2.4 education +- → 2.2 search knowledge +- → 5.1.1 action – 5.1.7 resource type*). It returns 19 hits, but the searcher does not look at any of the entries in the result list. Instead she enters a new query (*5.1.1 action – 5.1.7 resource type*), which start transition S8_T14.

In all these examples the searchers show that their query formulation is limited independent of the resource type they interact with. It is noticeable that we only have one example of a searcher who consciously becomes aware of this problem and manages to break out of her habit in order to perform Boolean AND-queries. Below we show excerpts from the point where the searcher changes her way of querying.

Example from Session 6:

> The searcher next (still in S6_S17) enters the query "henriksen vera" (*2.4 education +- → 2.2 search knowledge +- → 5.2.1 action – 5.2.9 resource type*). She decides to start exploring the result list (*5.2.1 action – 5.2.9 resource type*) by looking at an entry under the heading "Kultur". This strategy turns out to be a success since the destination of the link is a review of one of Henriksen's books (*5.2.6 relevance judgement*). The next four pages she examines are however not relevant. The searcher understands that she has entered a Boolean OR-query, and that she should "find a way to combine [the query terms] so that the hits returned are only about Vera Henriksen". She wonders whether there is any search help (*5.2.9 resource type*) that might tell her about the query syntax and returns to the query form. There is unfortunately no help to find so the searcher enters the (correct) Boolean AND-query "henriksen og vera" (*2.2 search knowledge + → 5.2.1 action*). The query results returned are all algorithmically relevant (*5.2.7 relevance level*) although the searcher at one point utters that they are not. This may be due to *2.7 uncertainty/affection,* and lack of *2.8 attention - →*; she does not understand that she has performed a logically correct query. She re-finds the book review and actually bookmarks it once more, before she re-enters the query (*5.2.1 action*) "henriksen vera". The searcher looks at two non-relevant articles and decides that it is better to state the query with an "and". The searcher considers herself to be done with Henriksen and backtracks (*5.2.1 action*) to the query form.

The searcher continues using the correct syntax for two more queries in this resource, but unfortunately does not transfer her "new" knowledge to later situations and transitions where she would have benefited from it.

In our view it would be fruitful to perform a similar analysis of search knowledge influence in a controlled setting. The findings are quite interesting and one research design for future studies could be to compare searchers with an LIS background and searchers who have not been trained to be experts in using bibliographical databases.

5.5.6 Influence of work task stages on WIS processes?

The nine WIS processes we have recorded can be categorised according to at what stage in the work task process they took place. None of the WIS processes were performed towards the end of the work task, but still a majority of the WIS processes were

concerned with the collection of clearly defined material (author information), which is comparable to what Kuhlthau (1991; 1993) calls the "collection" stage and which usually is the last stage before the completion of the task. We believe that this has to do with the special characteristic of the particular work task generating these five search processes (S3-S6 and S8). In addition S1 had a similar search task goal; to collect bibliographic and biographic information on a specific author. We have categorised the nine sessions into two different stages; the author data collection sessions belong to the collection stage while the remaining three sessions are in the pre-focus formulation group. We have then examined the data stemming from our analysis of the other empirical research questions to see whether any particular patterns turn up. The most important differences between the two groups are listed here:

- The WIS processes belonging to the pre-focus group are characterised by consisting of comparatively fewer situations and transitions than the collection group;
- The situations taking place in the WIS processes belonging to the pre-focus group are quite long compared to those generated by the collection group;
- Transitions last longer in the WIS processes generated by the pre-focus formulation group;
- Search knowledge is seldom identified as an influence on other attributes during WIS processes belonging to the pre-focus group, while it seems to be an important influence during "collection processes". This is probably due to the special characteristics of the work task and search task.

These findings can be used to claim that the searchers who perform the pre-focus WIS processes are inclined to be more thorough during the data collection. They are, we believe, still in the process of formulating the goal of their work task and therefore spend more time during both situations and transitions to "look around", in order to get ideas and inspiration to be used to shape their work task. The search task strategies used by the pre-focus group also seem to be less targeted, in the sense that these searchers have not to such a large degree decided which resources to use before starting the process.

5.6 Summary of the empirical study

We believe that the study presented in this chapter is a convincing argument for the usefulness of our method schema. We have shown how an explicit method can be used to collect and analyse web search behaviour generated by real work tasks.

The method can be used to provide knowledge about how people behave when they use the Web in order to solve larger work tasks and it can be used to analyse why they behave in that way.

The empirical data used in our study is limited with respect to the kinds of processes it covers and can only be used to describe and analyse the WIS processes of the participating searchers. On the other hand the data contains a large number of situations and transitions, opening up for detailed analysis.

The analysis made it possible to divide the search sessions into two groups, reflecting the stage of the work task that generated the processes. This indicates that the *work task stage* plays a direct role for the execution of web information search processes for these searchers. In addition we found, through micro-analyses of the WIS processes, that the *work task goal* directly influences the *relevance level* used when searchers evaluate the resources. Indirectly WT goals influence the amount of *time and effort* spent to perform the relevance judgement, because topical relevance judgements are faster and easier to make than situational. These findings support the argument that the work task directly influences WIS processes and it suggests that we should take the characteristics of the work task into consideration when analysing web searching. In other words the work task should not be replaced by the search task when analysing information behaviour on the World Wide Web. WIS processes being an example of information search behaviour also means that our findings support Vakkari's (Vakkari & Hakala, 2000; Vakkari, 2001) claim that work tasks directly affect information searching.

With respect to the work task-search task relationship we see that one work task (the creation of an author web resource) has generated 7 of the sessions we have studied. These sessions do, however, represent different kinds of search tasks. 5 sessions handle search tasks that closely resemble those performed by intermediaries. The other sessions generated by this work task have to do with the structuring and design of the Web resource. The searchers' evaluations of resources during the WIS processes were obviously influenced by the search task types, which in our analysis can be seen in the relationship between work task goal and relevance level. We have therefore suggested that further research should be invested in examining the relationship between the relevance level and the goals of work task and search task respectively.

The significance of dividing WIS processes into situations and transitions is recognised when we look at the time spent in situations/transitions. Searchers spend more time

examining resources of possibly direct use for them than they do finding such resources via search tools or other meta-information resources.

It is also interesting to look at the number of situations and transitions needed at different stages in the work task, as well as at the length of situations/transitions at different work task stages. It can be argued that this stems from the exploratory needs of searchers in pre-focus WIS processes, but this needs to be supplemented with more varied data to make a general claim.

Situations and transitions seem to be quite similar, differing only with respect to the relevance-attributes. Since we have not made an exhaustive list of attributes, we can only point out some possible differences that the method's domain does not explicitly cover. The action-attribute can be given the values "linking", "searching", and "scanning". We used this scheme in Pharo (1999) to characterise the "action"-dimension in our Web information search strategy-model (see Subsection 4.5.7). The following hypotheses might be worth testing: 1) during search situations searchers mainly scan pages; 2) during search transitions searchers mainly follow links and enter queries. The "resource type" is another attribute that can be classified in order to identify how they affect situations and transitions, since it is evident that meta-information resources would be those primarily used during transitions. We have shown that the amounts of time spent during situations and transitions differ and the same may also be true of the effort invested.

We have not examined whether the effects of other categories on situation attributes are different from their effects on transition attributes. Such analysis might be possible by developing classes for the individual attributes, which would open up for a more fine-grained analysis of relationships.

Situations are believed to be richer in variety compared to transitions, meaning that the searchers are inclined to be more creative and imaginative when involved in situations. This in turn may cause searchers to invest more time and effort as well as activate their information spaces more often and use the accumulated results to a larger degree. When studying the result of the analyses dealing with our empirical research questions, it is not possible to identify what, if any, situation specific attributes in particular that affect other attributes or are affected by other attributes during situations. The relevance attributes being the only exceptions. We have identified that the frequency of situation and transition attributes reflects the time spent during situations and transitions. The analyses have, however, not reflected a similar pattern; with the distinct exception of the WT goal, transition attributes are relatively more often affected by the attributes we have examined

the effect of. In other words it seems that ST strategies, work task knowledge, and search knowledge play more important roles during transitions than situations. One explanation may be that situations demand more from the searchers than transitions, which means that it may be harder to get them to voice their thoughts during situations. Thus some aspects of situations may have escaped our method. Additional studies are therefore needed to find out more about other factors that affect situation specific attributes.

In the next, and final, chapter we will discuss our method schema's virtues, make our conclusions, and point out some directions for future research.

6 Summary and conclusions

This chapter serves several purposes; it provides a summary of the foregoing chapters; it discusses implications and limitations of the work presented; and it examines how further research on Web information searching can be performed using the SST model.

In order to ease the reading of the summary and to make our perspective clear to the reader we start by repeating the goal of the project and clarifying the terminology used.

6.1 Goal and terminology

The purpose of this study has been to develop a method schema for the analysis and description of work task-based Web information search (WIS) processes.

A *method* is, according to Bunge (1967, p. 8), a procedure for handling a set of problems. We have used Newell's (1969) specifications to further clarify the nature of a method. According to Newell a method consists of the following three parts: (1) a problem statement or domain, (2) a procedure – often seen as the "method" in conventional meaning, and (3) a justification. The domain is used to represent the problem, for example in the form of a conceptual framework stating the properties of the problem and their relationships. The procedure is "what delivers the solution to the problem" (Newell 1969, p. 370), i.e. it specifies the rules for data collection and analysis. The justification provides the arguments in order to make it rational to believe that the method works, for example by providing a case representing the domain and using the procedure to show that it can deliver relevant findings.

A *method schema* is in the format of a method, only at a higher level of abstraction than a method (Eloranta, 1979). This means that it can leave one or more aspects of the method uninterpreted, represented only through their plain names, and some aspects of the method may be left out (even lacking naming). In other words a method schema is an abstract representation of one or more methods – a generic model.

In the format of a method our research problem can be presented the following way:

The *domain* of work task based WIS processes should be expressed in the form of a conceptual framework. It should take into consideration the main categories of such processes and present them in a comprehensible way. The framework should 1) be suitable for analysing search processes in the Web, 2) provide an effective level of

abstraction, 3) suggest attributes or dimensions of the categories that capture relevant features for understanding and explanation of their interplay, and 4) suggest classification of the categories or their attributes that allow for asking fruitful questions or detecting any systematic connections between them.

The *procedures* should include operations that are necessary to perform in order to analyse WIS processes and the factors that affect them. This involves procedures for collecting data that represent characteristics of WIS processes and affecting factors, also procedures for analysing data to discover the interplay in WIS processes that occur both between the categories as well as within them.

The *justification* of the method schema is twofold; it can be justified by showing its ability to solve designated research problems of the domain, i.e. by using it to generate a method for describing actual WIS processes. In addition the method schema's domain should be assessed using the formal requirements put forward by Bunge (1967) and by comparing it with previous attempts at modelling information behaviour (refer to Subsection 3.1.1).

6.2 Summary

We will now give a summary of the dissertation and specifically look at how the research problem has been treated in the foregoing chapters. We will refer to how the individual chapters relate to the method schema's individual parts.

In Chapter 2 we discussed the characteristics of the World Wide Web and what previous studies have been able to reveal about Web searching and Web search behaviour. The mapping of the area made it clear that there are still many research problems left in order to understand what happens during WIS processes. The complexity of the research area suggested that it is necessary to develop a domain description in the form of a conceptual framework, which additionally seemed to be a suitable format for pointing out fruitful research problems. The strategy of disjointed incrementalism (Braybrook & Lindblom, 1963) was used to describe some important features that seem to characterise web searching, an approach that helped us in coming to terms with some of the how's and why's of WIS processes; such as the limited number of often less than optimal strategies used by searchers; their adjustments of goals during interaction; and the role of the work task process seen over time. We ended the chapter by pointing out some strategies for developing the method schema.

Chapter 3 reviewed previous models of information behaviour in order to see what input they could give to our domain description. The models that we evaluated were collected from both the information seeking as well as the information searching/user centred IR literature. The criteria used for evaluation were partly content-based, i.e. focusing on our research problem – how they could deal with work tasks generating Web information search processes – and partly they consisted of formal criteria by Bunge (1967) for assessing theories. Although the models did not claim to be theories of information behaviour, they nevertheless represented attempts at mapping the research area as well as pointing out ways to analyse information behaviour. Bunge's criteria therefore were judged to be relevant for our purposes; focusing on epistemological as well as methodological aspects of theories. The evaluation showed that none of the existing models could satisfy our content-based criteria, since most of the models were too general in that respect. We did however find that there are elements from the two traditions of information behaviour research that could be combined, but that the existing models are firmly grounded in one of the two traditions. When examining the models in light of the formal criteria we found that they differed very much. Some models only satisfy criteria dealing with the guidance of research, while others only present a map of (parts of) reality. The most relevant model for our purposes appears to be Ingwersen's (1996) cognitive model of IR interaction. It combines features of user centred IR and information seeking, but it is not explicit enough for our purposes. In addition Belkin and his colleagues' (Belkin, Marchetti & Cool, 1993) conceptualisation of information searching as a series of episodes also seemed fruitful. There are however characteristics in all models that can be used to represent our domain and Table 3.1 sums up important characteristics of the previous models. In Table 3.2 we present the assessment of the existing models using the formal criteria.

One might expect to find input to our schema's procedures (the "methods") in the documentation of the previous models. To a certain degree we were able to find some data collection techniques used by these researchers. We discuss these and other procedures found in the information behaviour literature as well as in the general research literature in Chapter 4. In Chapter 4 we also categorise important characteristics found in the information behaviour literature into categories and attributes that constitute a sketch for our domain. A pilot study of three searchers at Oslo University College's Faculty of Journalism, Library and Information Science was performed which served three purposes: it combined a variety of different collection techniques, thus showing the applicability of parts of a procedure. It verified the importance of the suggested categories and attributes. Finally, it applied grounded theory to identify important attributes that we had not identified in the literature. The chapter presents the method schema's domain and

procedure and discusses a proposed set of attributes for the domain. In addition it discusses how procedures could be related to qualitative and quantitative research approaches. The chapter presents the proposed search situation transition (SST) method schema for work task based WIS research. Its contributions lie in: 1) suggesting a strategy for organising WIS data into search situations and transitions, 2) proposing categories and attributes that play essential roles in work task based WIS processes, and 3) suggesting a strategy for discovering the interplay of categories and attributes in the situations and transitions. Finally the chapter compares the SST method schema and previous models with respect to their ability to meet the formal criteria of Bunge.

Chapter 5 is dedicated to the justification of the method schema by creating a specific method for the analysis of the WIS processes performed by a group of LIS students working on their theses. In this empirical study we investigate the effects on the WIS process of work task goals and stages, the search task strategies, searchers' task knowledge, and searchers' search knowledge. We find that the method is suitable for analysing such effects and conclude that our method schema can be justified as a tool for analysing task based WIS processes. We will, in the following sections, point out the unique features of the SST model and discuss possible uses of the schema. We believe that our model conceptualises the joint areas of information seeking and information searching in a way that is an important step forward for research in this area.

6.3 The need for methods for information behaviour research

The traditional meaning of methods has focused on what we have termed procedures. This refers both to methods of collection and analysis. Our perspective is broader; focusing on how to reveal problems related to web information search processes. We have chosen our methodological perspective because it offers a fruitful way to treat the area of information behaviour. Information behaviour is an example of a complex and ill-structured problem, which can best be described or conceptualised in the form of a domain description. Our methodological viewpoint explicates what is the description and what are the possible techniques (the procedure) to be prescribed to treat the problem, for example to reveal hidden relationships or entities in the domain.

Complex and ill-structured problem areas can be treated with a variety of methods; the domain description may differ depending on what issues of the problem one wishes to examine. Measuring their ability to deliver solutions can compare the power of methods. Newell (1969) points out three dimensions, along which this ability can be measured.

These are the probability of solution, the quality of solution, and the amount of resources needed.

Since "information behaviour" is such a complex problem it would be inadequate to create an exhaustive domain description, as well as prescribe all possible procedures that could be used to deal with the problem. WIS processes represent a sub-set of information behaviour, but still there is a multitude of factors that may affect the processes. The method schema is a suitable format for problems of this kind, since it allows leaving less relevant entities out of the method in favour of specifying appropriate methods when particular problems arise. It is necessary to treat such problems in parts.

The method schema provides a framework that can be used to create more powerful methods than the power of a single general method. A method designed for a narrow purpose increases the probability of solution, since the method is not cluttered with superfluous elements in its domain or procedure, which alternatively might have been a source of confusion and thus increased erroneous use of the method. Additionally such a method does not remain on too high a summary level. Also such a specific method would probably strengthen the quality of solutions, given that it is designed for the particular problem rather than a group of problems. With respect to the third dimension of power, costs, it can be harder to decide what requires the most of resources; using a harder-to-use general method or developing and using a specific method for each problem. We believe, nevertheless, that the weight of the first two dimensions favours the use of specific methods designed according to the recommendations of the SST method schema. However, the use of qualitative video data analysis, as in our case, is always labour-intensive and the effort must be justified by the understanding attained.

We have not found other tools treating information behaviour in a methodologically similar way, thus the SST method schema could be said to be a pioneer work in the LIS field. We believe that the method schema can be applied in such a way as to focus on a range of issues dealing with information behaviour on the Web; we have demonstrated this by developing a method that focuses on a particular group of searchers working with a specific type of work tasks. For further discussions on the justification of the method schema see below.

6.4 Justifying the SST method schema

The third part of a method is its justification, without which what is proposed in the domain and procedure becomes speculation. We chose to perform a case study of a

group of students using the Web in order to help them in solving a particular kind of work task.

In the case of empirical social science problems one cannot formally prove that an algorithm solves a problem. One has to contend rather with an indicative proof in the form of one or more representative examples. The most basic case, as in the present project, is to present *one* representative case to informally prove that the method generated by the method schema works at least in this case. It then remains a matter of argumentation to convince the reader that the method (schema) is extendable to other similar research problems.

We used our method schema to develop a method for analysing the students' WIS processes. Some empirical research questions concerning the influence of work task, search task, searchers' work task knowledge, and their search knowledge were formulated. We selected attributes of searchers' information behaviour that have been identified in previous research as playing significant roles for our domain description. The procedures used in collecting and analysing the data were also well known from previous research in information behaviour.

We have proposed a set of categories and attributes and suggested their relationships. In our view the use of various procedures for data collection and the way we have created stories secure a rich representation of what happens during WIS processes. By using a combination of procedures for data collection we believe that the case represent valid data. The combination of the different data into stories makes it possible to combine them to identify relationships that not necessarily would have been identified if the individual data sets were analysed separately. Our procedure is flexible but we recommend the use of triangulation to secure the validity of the data that are collected. During the analysis it is recommended that a guide for data classification is developed to secure analysis consistency and reliability.

Understanding small adjustments and changes in the search process, as a result of how searchers interpret the results of their interaction with the Web, has been important. Clearly in accordance with the strategy of disjointed incrementalism (DI) (Braybrook & Lindblom, 1963), originally developed to explain decision making, we have identified that small steps in the WIS process, i.e. single situations and transitions, may slightly change the direction of the process. Our analysis showed that the searchers adjust their work task goals as a result of the information found and not found. This also will influence their search task goal and search task strategies. In other words the decisions about how to

continue the process to a large degree depends on the information acquired by the searcher throughout the process, not on a rational strategy taking into consideration that the searcher would need to have total information about all factors influencing her problem. The searcher, being constrained by a set of factors, for example her environment, takes into consideration the resources she is able to find and adjusts her task to these. We believe the format of the domain description is fit for such purposes, which makes it possible to explicate what categories and attributes relate to one another, and what effects the relationships trigger. We have showed that the method is able to identify such features in WIS processes thus indicating that the method schema is usable for this kind of analyses.

The justification of the method used in the empirical study could be divided into two, which means 1) that if we were able to use the method in analysing a case representing typical web information search behaviour, and 2) if the results of our analysis seemed reasonable, then this would provide a convincing argument for the method's usability for analysing WIS processes.

In our opinion the search sessions analysed in Chapter 5 represent typical search behaviour on the Web: 1) The searchers use the Web in order to find information that may be of help to them in solving real work tasks. 2) The searchers did not represent a group of Web search experts, although they had a few years of experience in using the Internet. 3) Their searching was targeted at the work task and search task in hand. They wanted to find a sufficient amount of information to execute their tasks in a satisfactory way. Whether the conclusions are convincing should be up to the reader to judge, but they indicate that it is worth the effort to take work tasks into consideration, when looking for influences on information behaviour, and our data support Vakkari's (2001) studies in the sense that the searchers' work task stage directly influences search behaviour.

6.5 Describing and prescribing information behaviour

During Web information search processes searchers are influenced by many different things. We have sought to develop a tool for learning about *real* human information behaviour rather than pointing out how it should ideally be.

This relates to the strategy of disjointed incrementalism (Braybrook & Lindblom, 1963). The strategy can be contrasted with rational theories of problem solving/decision-making demanding that those responsible for executing a task or solving a problem need to have

total information about all factors influencing the problem and all consequences of possible solutions in order to choose the optimal solution.

Most searchers, however, are not interested in developing their searching skills to the utmost. They only want to be able to cope with their information needs. Information searching is not their focus; they want to be able to perform their tasks. For this reason information behaviour, and WIS processes, may appear quite haphazard.
Web searchers do not have total control over the WIS processes; they react to the resources they interact with by sharpening their search task strategies, they reformulate their work task and search task goals, they are hindered by lack of appropriate knowledge about how to squeeze the needed information out of the resource etc. Yang (1997) has observed something similar when studying the search behaviour of students using the Perseus hypertext system. She says, "the study also demonstrated that the learners did not follow a prescriptive, predetermined plan." (Yang, 1997, p. 89), and continues to argue, referring in her paper to Suchman (1987), Schön (1990), and Norman (1988) "that purposeful behavior is ad-hoc or extemporaneous, rather than deterministic and logical." (Yang, 1997, p. 89). Also Fidel and her colleagues (1999) have noted the opportunistic behaviour of Web searchers, finding among other things that high school "students did not have much tolerance for long lists on the results page".

Knowledge about real information behaviour on the Web can supplement studies that focus on query formulation and retrieval efficiency of various search engines. We must learn how people attack work tasks and search tasks with the help of the Web in order to understand why they choose specific search engines or subject indices, and we should become familiar with the various factors that make searchers reformulate and/or abandon their use of search engines. From our perspective it is not enough to learn what are the most efficient ways to use Google or Alta Vista. We strongly emphasise the value of coupling traditional IR research on search engines efficiency and usability with knowledge about searchers' expectations and the kind of information needs they have when they choose to use such services.

6.6 The flexibility of a method schema

A method schema is characterised by being more flexible than a single method. It should provide a framework for describing the problem domain and suggest relevant procedures for analysing the problem. The domain description should represent important categories at a level of granularity which is meaningful for exploring research questions in that area. On the other hand the method schema is more general in the sense that it can leave

elements uninterpreted and choose to represent them only through their names. It can also leave elements of the methods out. This does not mean that the method schema is less meaningful than a method; the schema should be presented in a way fostering easy adaptation of new elements to generate a particular method, for example to include uncovered categories and attributes in a new method's domain.

The choice of focusing on a method *schema* was based on three observations. The first of these was the recognition of the Web as a multi-faceted information system providing a large variety of interaction methods and resource types. The Web can be used as a gateway to a lot of information systems, such as various databases, that previously had to be accessed separately via their own client programs. The Web also provides a human-computer interface that many of these service providers have adapted to, making the interaction methods more standardised – at the cost of some advanced query options that have not been implemented in the web interface. On the whole a lot of standardisation has taken place in the Web, for example concerning the syntax of search engines. Nevertheless the multitude of different information systems and sources available represents one dimension of the Web environment that necessitates the flexibility of a method schema for analysing WIS processes. Having a *method schema* rather than a single method makes it possible to generate methods to study, for example, the relationships between specific instances of resource types and interaction techniques and how these affect or are affected by other factors.

The second reason for selecting to construct a method schema was the fact that the Web is available to a large variety of populations. Thus we need to take into consideration that the method should be usable for describing and analysing user groups stemming from different kinds of environments and varying in demographic as well as other factors. Not so long ago advanced information systems were the domain of a more homogeneous user group; they were highly trained users with a high level of search knowledge and/or domain knowledge. We need to take into account that this no longer is the case.

Thirdly the Web can be used for a large number of reasons and the diversity of the work tasks that generate web use necessitates a flexible tool for taking into consideration the specific factors that seem to be dominant for the work task type in hand.

We are convinced that the level of generality we have used to develop our tool is fruitful; it is aimed at taking into consideration some major categories that influence WIS processes, as well as suggesting some important characteristics or attributes of these categories. A general method for analysing WIS processes risks, in our view, two fates. It

may bring too many categories, attributes, and relationships into consideration in the domain description, which may unnecessarily complicate the handling of the problem. Or, on the other hand, it can become too general and leave out significant categories etc.; the analysis would then run the risk of becoming trivial and not worth the effort. The flexibility of a schema facilitates the generation of methods to deal with problems of different generality.

6.7 The SST model and qualitative/quantitative research

The method schema provides a tool with a rich set of opportunities to be used for the analysis of WIS processes. In Chapter 5 we justified our method schema by showing how it can be used to analyse the WIS processes of a small group of searchers with quite complex work tasks; this exemplified a qualitative approach to the problem. It also exemplified how a variety of data collection methods can be used to collect a sufficiently rich set of data. In our study we wished to identify as rich an interplay of categories and attributes as possible to show the SST method schema's applicability.

The method schema is flexible and provides a way to generate methods to treat research problems of different complexity. Qualitative research designs would be necessary to deal with the more complex research problems.

The formulation of hypotheses in the form "attribute X leads to an effect on attribute Y under such and such conditions" would facilitate a quantitative research design. To provide as controlled a setting as possible, Borlund's (2000) method for evaluating interactive IR system could be used. Borlund has shown that there is no significant difference in how searchers perform when they treat search tasks based on their own information need and when their points of departure are given as simulated work task situations[22]. Borlund recommends that the simulated work task situations should be tailored to the searchers, i.e. to create situations that are within their domain of interest. Although the method was developed for other purposes, we believe that simulated work task situations can also be used as a point of departure for doing analyses of WIS processes. Whether simulated work task can be used to analyse how people search the Web when performing very complex tasks is not clear. We believe that it can be difficult to *stimulate* searchers to be as involved in very complex simulated work tasks as in real ones. Thus it may be difficult to identify how searcher attributes like "motivation",

[22] "A simulated work task situation is a short 'cover-story' which describes a situation that may lead to IR and seeking." (Borlund, 2000, p. 80)

"tenacity", "uncertainty", and "attention" influence WIS processes without analysing real work tasks.

When one wishes to study the interplay or effects of only a few attributes, a quantitative research approach would be fruitful. The searcher group that one wishes to examine could be tuned by focusing on searchers with similar education, age, search experience, work experience, and so on. These attributes could then be omitted from the analysis since they are constant, and one might choose to look at how other attributes generated effects on the WIS process. The use of simulated work tasks provides an opportunity to keep the necessary stability.

Since the complexity of the method used stems from the problem that one wants to examine, we shall now discuss some problem types that the method can be used to investigate.

6.8 Problem types for which to use the SST method schema

In Chapter 5 we used a method to analyse work task based WIS processes by a small user group, in order to identify the role of, among other things, work task goals and the work task stage of the searchers. The purpose of that study was to justify the method schema.

Future use will show whether the schema can also be used as a tool for applied research, by which we mean research performed with the intention of having a value for society, i.e. research performed to have practical applicability (Hellevik, 1999).

Two areas are obvious candidates for such treatment. The first group of problems focuses on system design. This includes issues dealing with the design of client program issues as well as those having to do with the structuring of Web resources. These are issues often treated under the "usability" label (Shneiderman, 1998; Nielsen, 2000).

In the other group we find those problems that are concerned with the searchers. Some examples are issues related to user training, problems dealing with Web source selection and Web source criticism, and the need to know how to develop dedicated resources for particular groups of searchers. The latter example is clearly related to both the user and system side of interaction. We shall sketch some possible research problems for each of the areas, starting with those related to the system side of interaction.

6.8.1 System-focus

It seems reasonable to believe that more knowledge about how attributes of the searcher influence the WIS process could help client program providers to tailor more customised software. Also knowledge about search tasks, and possibly work tasks, may make it easier for software and web resource providers to facilitate predictable interaction.

To focus on such issues it will be useful to start with a hypothesis on what relationships to investigate. Some good candidates are search knowledge and task knowledge, which we know from previous studies (Marchionini, Lin & Dwiggins, 1990; Marchionini et al, 1993; Hsieh-Yee, 1993; Hölscher & Strube, 2000) help searchers to construct effective strategies. Too little is known about the exact effect search and task knowledge have on the ingredients of WIS processes and we suggest comparing searcher groups with different levels of expertise to see how this affects the process. Such a study may well benefit from using simulated work task situations, as discussed above.

Search process attributes that belong to the system side of interaction should be examined with particular attention. In the method used in Chapter 5 those would have been "action", "resource type", and "technical problems". Methods generated from the method schema could be used to focus on the categories and attributes that cause effects on such "system attributes". This in turn could be used to enhance our knowledge and help designers to create more "intelligent" agents and contextual help systems. We believe that it would be particularly interesting to learn whether explicit process attributes such as "time", "accumulated results", and "accumulated effort" systematically influence the system attributes, and, if so, whether this is dependent on characteristics of the searcher.

In "system" we also include web pages and resources, the design of which might benefit from taking into account results of analyses using our model. We may, for example, develop a classification scheme for the "resource type" and study how different resource types are connected to other attributes. In this way we may identify the advantages and limitations of using different kinds of resources for different kinds of work tasks and search tasks.

6.8.2 Searcher-focus

Seen from the searcher side of the interaction, methods that can be generated using the schema may increase knowledge about how to train searchers in using the Web more efficiently. Another perspective could be to make searchers more aware of the limitations

of WIS processes, for example by performing similar analyses of different kinds of information systems to compare their strengths and weaknesses.

Local studies at Oslo University College (Fagerli, 2000) of the use of quality-controlled electronic sources have shown that these services tend be less used in favour of the use of general Web search tools. At the College the use of database hosts such as ISI and Ebsco, which provide access to databases such as the Science Citation Index, Medline, and ERIC has decreased. This may be the effect of a growing notion that "everything" can be found in the non-restricted areas of the Web.

The study performed in Chapter 5 revealed that the LIS-students preferred to use a query-syntax that has been developed for interaction with bibliographical databases also when they used other kinds of databases. Such default "erroneous" use of search knowledge, or similar professional habits, may be characteristic of other searchers interacting with Web and our method schema can be used to identify such erroneous behaviour and make searchers aware of its consequences. In our case, however, this finding was not expected and therefore it may be necessary to specify hypotheses suggesting a relationship between particular searcher attributes and attributes characterising the WIS process. For example, one could look at searchers who are trained journalists and see how this influences their "source criticism" (a work task attribute) during WIS processes.

We may develop methods that take into account how a particular group of searchers deals with various work task types, how different searcher groups would attack similar work tasks, the role of social environments etc. We shall examine a few research problems that it would be interesting to follow up and point out how methods to investigate them could be based on the SST method schema.

One research problem that needs to be more thoroughly investigated is how work task goals influence WIS processes. We have found traces of the effect of work task goals in our case study. Further research in this area could be done by giving assignments to students and following them in a similar way to what was reported in our study. The work task could be integrated as part of their training and one might give the students three different work tasks, demanding similar workload, distributed so that one might compare the work tasks. Then it would be possible to take into account factors such as Web information search training (search knowledge), expecting this to increase from the first to the third task, and searchers' experience.

A close follow-up of the searcher would make it possible to investigate the work task-search task relationship. This would be interesting to study in order to learn about the characteristics of search tasks that searchers wish to execute using the Web and to compare them to those that they would choose to solve via other information systems or sources. Similar search task types could then be compared for different searchers.

Another research problem that can be investigated using our method schema is the interplay between organisational search policies (strategies) and WIS processes. This might focus on how organisations utilise Web as an information search source, necessitating the use of a wide range of procedures. Recordings of WIS processes would need to be supplemented by interview data of managers, organisations' information specialists, consultants performing relevant work tasks etc.

We have proposed some research problems to emphasise the wide range of research designs that the SST model facilitates. In the next subsection we shall discuss other uses of the method schema.

6.9 The SST model applied on other information systems

Our review of existing models of information seeking and information searching and retrieval showed that there has been an increased interest in the human factor since the end of the 70's. Previous models have, however, only to a certain extent been able to take into consideration a person's problem solving, or information seeking, characteristics, and combined them with factors that play a role during searcher-information system interaction. The domain of the SST method schema introduces explicit relationships between categories that represent work tasks as well as information seeking and information searching characteristics, to an extent that is rare in the literature.

Although our method schema has been developed to deal with information searching on the World Wide Web it may be suitably employed to analyse and describe the uses of other information systems. The domain has no explicit limitations hindering the description of task based information search processes in other systems, and the data collection procedures that we have suggested have been used in previous studies of information searching and seeking behaviour. The search process category with its focus on situations and transitions is organised having the complexity of the Web in mind and is developed taking into consideration that switches between information sources take place within the same client program. Transitions can, however, be compared to what happens when a person with an information need decides to stop using an academic

journal in order to find whatever she is looking for in favour of asking a colleague. The transition for getting hold of that colleague could involve finding a phone book, looking up the person's number, and making a call, or the simple action of knocking on the door of one's next door neighbour. The interaction with a library catalogue to find the shelf number of a needed book can also be considered a transition, while the collection of, and interaction with, that book would constitute the situation part of the process.

We have suggested a set of attributes to describe transitions and situations, but we have not stated that any of these attributes are mandatory. On the contrary we have chosen a method schema because it allows leaving aspects of the method uninterpreted as well as leaving out some aspects of the method (Eloranta, 1979). This implies that it is possible to add attributes that are characteristic of the searcher-information system interaction that one wants to analyse.

6.10 Information seeking and searching – a need to combine efforts

Most researchers who have tried to develop general models that attempt to conceptualise both information seeking and searching and retrieval have had their point of departure in the information seeking tradition (e.g. Wilson 1996). The present work goes further by integrating characteristics of the searcher, her surroundings, the work task, the search task, and the interaction process between searcher and search system. Seldom have characteristics of the actual searcher-system interaction been coupled with categories such as the searcher's work task. However, in a recent study Vakkari has, among other things, found that work task stages influence the number of query terms used as well as the search tactics used in different stages (Vakkari, 2000).

We have claimed that the schema's domain is one of the first attempts at creating a conceptual framework using characteristics from the information seeking tradition as well as the user centred IR/information searching research domain. Ingwersen's (1992; 1996) cognitive model of IR interaction also included elements from both traditions, and Vakkari's recent theory of the task-based IR process (Vakkari, 2001) focuses on how the work task stage influences the information search process. Of these two, Ingwersen's model is the most holistic and we should like to point out some of the extension made by the SST method schema over his model.

The SST method schema explicates the process view on information search behaviour, which was only implicit in Ingwersen's model. In addition we go into more detail in organising the context in which information searching takes place, in particular by

organising work task and search task into categories of their own. Search task was not treated explicitly by Ingwersen and work task was seen as an attribute of the searcher's cognitive model. Our method schema also is clearer in pointing out the relationships that directly and indirectly affect the search process. On the other hand Ingwersen also focuses on aspects of the IR system setting, such as search language and indexing rules. In this respect his model goes deeper in modelling some elements in the search process.

We believe that to combine the efforts from both traditions dealing with information behaviour is a necessary step forward for our research area. It will be useful to learn more about the context of people's actual interaction with information systems for several reasons. When discussing problem types that the SST model could be used to solve we pointed out the need to improve information systems and user training, but there is of course also the very important and continuous need to expand our general knowledge in this area. We believe that the isolated attempts at attacking the problems from one tradition or the other should be reduced in favour of research that combines seeking and searching issues.

In particular it would be fruitful if researchers from both traditions could combine their effort to participate in common projects. It would be increasingly interesting if researchers from the system-oriented side of information retrieval could join such projects. The number of researchers that place themselves within the cognitive viewpoint of IR (Ingwersen, 1999) signifies that such a direction is emerging. Both Ingwersen and Belkin, who are probably the two most influential researchers in that tradition, have tried to develop models of searcher-intermediary-system interaction. The Monstrat (Belkin et al, 1987) and Mediator (Ingwersen, 1992) models aim at taking into consideration characteristics of the searcher as well as the problems that generated the information need. The models were briefly discussed in Subsection 3.1.3 in the discussion of Ingwersen cognitive IR interaction model. These models were developed during the "pre-Web period" and were targeted at developing dedicated interfaces for special types of databases. The models aimed at describing what features should be part of information systems' interfaces in order to make the searcher give as accurate a description of her information needs as possible. It has not been our intent to develop new features of information systems. The use of our schema may, nevertheless, lead to the identification of relationships that focus on system characteristics not specified in the Monstrat/Mediator models.

6.11 Limitations of this study

In this subsection we shall start by examining our point of departure for the study in order to see the effects of it being founded in the LIS domain. Then we shall discuss some specific features of the method schema and their limitations.

The point of departure for this study was to develop a method schema taking into account the role of work tasks' influence on information behaviour. When creating our schema we therefore focused on literature from the information behaviour tradition. Since the field seemed, and still seems, to be moving towards a consensus with respect to a common user-oriented perspective, we chose to limit ourselves to the literature within and related to this domain. We believed that this would lead us to a rich set of sources for characteristics of information behaviour and generate a sufficient amount of ideas to follow up both with respect to the development of the domain description and specific procedures ("methods") for collecting and analysing data.

A further step forward would be to integrate empirical findings from usability and human computer interaction (HCI) research (e.g. Nielsen, 1995; 2000; Shneiderman 1998). There are surprisingly few examples of co-operation between the researchers on information behaviour and usability, and it is also hard to find articles where information behaviour researchers take usability issues into consideration. The challenges of trying to integrate information seeking, searching, and retrieval issues were, however, sufficiently consuming, and something should remain for future research.

This means that we have not intended to take into account issues related to the graphic design of web pages and resources. Neither is our model focused on identifying how functionalities implemented in different kinds of web resources are met by the searchers' ability to use them. On a more general level, however, we may well classify our empirical study as an example of analysing Web usability.

If we look more specifically at our method schema, one central feature that needs to be more thoroughly investigated is the difference between situations and transitions in the WIS process. We have indicated that situations may be the more demanding of the two and that this may be a reason why we have not been able to spot exactly how they differ. In turn this indicates that the procedures used for collecting the relevant data have not been optimal. Talking aloud during the WIS processes, which requires extra effort from the searchers, may favour those parts in the processes that require the least from the searchers, i.e. the transitions. One possible way to avoid this potential problem would be to have searchers comment on their WIS processes immediately after the session is

finished. This would obviously be more time consuming. On the other hand, to get as rich an understanding as possible about the similarities and differences of situations and transitions is necessary in order to claim that this is a fruitful way of conceptualising WIS processes; the research area is complex enough without us having to further complicate things.

Related to this is the different nature of the attributes that we have presented. Some attributes are obviously easier to identify than others. "Actions" performed during the search process are directly observable, while both "remaining needs" and "information space" are attributes that need interpretation by the researcher. Very roughly the attributes can be categorised as being either external or internal. Many of the searcher attributes are clearly internal, for example "motivation", "tenacity", and "uncertainty". External attributes such as the searcher's "education" and attributes related to work task and search task can be identified more easily. For this reason there is a danger that attributes may be represented in a way which do not reflect their importance, even if one makes the searchers voice their thoughts and feelings during the search process.

6.12 Conclusions

The present study provides a framework in the form of a method schema that can be used to describe and analyse how work task-based web information search processes are influenced by a range of factors. The method schema is rooted in existing information behaviour literature and represents a new way of combining issues related to information seeking and information searching/retrieval. In this sense it is a pioneer work and one which may lead information behaviour research towards a more holistic and integrated perspective on issues having to do with all participants in the information search process.

References

Allen, T. J. (1969). Information needs and uses. *Annual Review of Information Science and Technology*, 4, 3-29.

Audunson, R. (1999). Between professional field norms and environmental change impetuses: A comparative study of change processes in public libraries. *Library & Information Science Research*, 21 (4), 523-552.

Bates, M (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13 (5), 407-424.

Belkin, N. (1990). The cognitive viewpoint in information science. *Journal of Information Science*, 16 (1), 11-15.

Belkin, N. J. et al. (1987). Distributed expert-based information-systems: an interdisciplinary approach. *Information Processing & Management*, 23 (5), 395-409.

Belkin, N. J., Cool, C., Stein A. & Thiel, U. (1995). Cases, scripts, and information-seeking strategies – on the design of interactive information-retrieval systems. *Expert Systems with Applications*, 9, 379-395.

Belkin, N. J., Marchetti, P. G. & Cool, C. (1993). BRAQUE: design for an interface to support user interaction in information retrieval. *Information Processing & Management*, 29 (3), 325-344.

Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982). ASK for information retrieval I: Background and theory. *Journal of Documentation*, 38 (2), 61-71.

Bergman, M. K. (2000). The Deep Web : surfacing hidden value [Online]. Available: http://128.121.227.57/download/deepwebwhitepaper.pdf [October 11, 2000].

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I. Cognitive, physical and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51 (7), 646-665.

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine: II. Cognitive and physical behaviors on research tasks. *Journal of the American Society for Information Science and Technology*, 52 (2), 118-136.

Borlund, P. (2000). *Evaluation of interactive information retrieval systems*. Åbo: Åbo Akademi [Doctoral dissertation].

Borlund, P. & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 53 (3), 225-250.

Braybrook, D & Lindblom, C. (1963). *A strategy of decision: policy evaluation as a social process*. New York: Free Press.

Bunge, M. (1967). *Scientific research*. Heidelberg: Springer-Verlag.

Byström, K (1999). *Task complexity, information types and information sources : examination of relationships*. Tampere: University of Tampere [Doctoral dissertation].

Byström, K. & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing and Management*, 31 (2), 191-213.

Cailliau, R. (1995). A Short History of the Web: Text of a speech delivered at the launching of the European branch of the W3 Consortium Paris, 2 November 1995 [Online]. Available: http://www.cs.ucy.ac.cy/courses/EPL422/readings/Cailliau-fra.html [February 18, 2002].

Catledge, L. D. & Pitkow, J. E. (1995). Characterising browsing strategies in the World-Wide Web. *Computer Networks and ISDN Systems*, 27 (6), 1065-1073

Chen, B., Wang, H., Proctor, R. W. & Salvendy, G. (1997). Design and use of browsers: a human-centered approach for designing World Wide Web browsers. *Behavior Research Methods, Instruments, & Computers*, 29 (2), 172-179.

Cockburn, A. & Jones, S. (1996). Which way now? Analysing and easing inadequacies in WWW navigation. *International Journal of Human-computer studies*, 45 (1), 105-129

Cohen, M. R. & Nagel, E. (1934). *An Introduction to Logic and Scientific Method*. London, Routledge.

Cosijn, E. & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management*, 36 (4), 533-550.

DCMI Type Vocabulary (2000) [Online]. Available: http://dublincore.org/documents/dcmi-type-vocabulary/ [March 18, 2002].

De Mey, M. (1982). *The cognitive paradigm*. Dordrecht, Reidel.

Denzin, N. K. & Lincoln, Y. S. (eds.) (1994). *Handbook of qualitative research*. Thousand Oaks, Calif., Sage.

Dervin, B. (1983). An overview of sense-making research: concepts, methods, and results to date. In *International Communication Association annual meeting, Chicago, May*.

Dervin, B. (1992). From the mind's eye of the 'user': the sense-making qualitative-quantitative methodology. In J. D. Glazier & R. R. Powelle (eds.), *Qualitative research in information management*. Englewood, Colo.: Libraries Unlimited, 61-84.

Dervin, B. (1999). On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing & Management*, 35 (6), 727-750.

Dervin, B. (2001). Sense-making studies [Online]. Available: http://communication.sbs.ohio-state.edu/sense-making/default.html [March 22, 2002].

Dervin, B & Dewdney, P. (1986) Neutral questioning: a new approach to the reference interview. *RQ*, 25 (4), 506-5l3.

Dervin, B. & Nilan, M. (1986). Information needs and uses. *Annual Review of Information Science and Technology*, 21, 3-33.

Dublin Core Metadata Element Set: Version 1.0: Reference Description (1998) [Online]. Available: http://dublincore.org/documents/1998/09/dces/ [February 21, 2002].

Dublin Core Metadata Initiative (2002) [Online]. Available: http://dublincore.org/ [February 21, 2002].

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation*, 45 (3), 171-212.

Ellis, D., Cox, D. & Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation*, 49 (4), 356-369.

Ellis, D. & Haugan, M (1997). Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation*, 53 (4), 384-403.

Eloranta, K. T. (1979). *Menetelmäeksperttiyden analyysi menetelmäkoulutuksen suunnittelun perustana* [The analysis of method expertise as a basis for planning methodology Education]. Tampere: Tampereen yliopiston.

Engelbart, D. (1962). *Augmenting human intellect: a conceptual framework*. Stanford Research Institute, Menlo Park, Calif.

Ericsson, K. A. & Simon, H. A. (1996). *Protocol analysis: verbal reports as data*. MIT Press, Cambridge, Mass.

Fagerli, H. M. (2002). *Personal communication*.

Fidel, R. (1985). Moves in online searching. *Online Review*, 9 (1), 61-74.

Fidel, R. (1993). Qualitative methods in information retrieval research. *Library and Information Science Research*, 15 (3), 219-247.

Fidel, R et al. (1999). A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society for Information Science*, 50 (1), 24-37.

Frohmann, B. (1992). The power of images: a discourse analysis of the cognitive viewpoint. *Journal of Documentation*, 48 (4), 365-386.

Frohmann, B. (1994). Discourse analysis as a research method in Library and Information Science. *Library and Information Science Research*, 16 (2), 119-138.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory*. New York: Aldine de Gruyter.

Gordon, M. & Pathak, P. (1999). Finding information on the World Wide Web: the retrieval effectiveness of search engines. *Information Processing & Management*, 35 (2), 141-180.

Ginman, M. (1987). *De intellektuella resurstransformationerna*. Åbo: Åbo Academy.

Gove, P. B. (ed.) (1971). *Webster's third new international Dictionary of the English language* (unabridged ed.). Springfield: Mass.: Merriam.

GVU Center's WWW User Surveys (2000) [Online]. Available: http://www.cc.gatech.edu/gvu/user_surveys/ [July 19, 2000].

Hansen, P. (1999). User interface design for IR interaction: a task-oriented approach. In T. Aparac, T. Saracevic, P. Ingwersen & P. Vakkari (eds.), *Digital libraries: interdisciplinary concepts, challenges and opportunities: proceedings of the Third International Conference on the Conceptions of the Library and Information Science (CoLIS 3)*. Zagreb: University of Zagreb, 191-205.

Hellevik, O. (1999). *Forskningsmetode i sosiologi og statsvitenskap*. Universitetsforlaget, Oslo.

Hewins, E. T. (1990). Information need and use studies. *Annual Review of Information Science and Technology*, 25, 145-172.

Hjørland, B. & Albrechtsen, H. (1995). Towards a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, 46 (6), 400-425.

Hof, R. D., Browder, S. & Elstrom, P. (1997). Internet communities. *Business Week*. May 5, 38-45

Hölscher, C. & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, 33 (1-6), 337-346.

Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Addison-Wesley, Reading, Mass.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science*, 44 (3), 161-174.

Hsieh-Yee, I. (1998). Search tactics of Web users in searching for texts, graphics, known items and subjects: a search simulation study. *Reference Librarian*, 60, 61-85.

Iivonen, M. (1995). Consistency in the selection of search concepts and search terms. *Information Processing & Management*, 31 (2), 173-190.

Ingwersen, P. (1982). Search procedures in the library analysed from the cognitive point of view. *Journal of Documentation*, 38 (3), 165-191.

Ingwersen, P. (1992). *Information retrieval interaction*. Taylor Graham, London.

Ingwersen, P. (1993). The cognitive viewpoint in IR. *Journal of Documentation*, 49 (1), 60-64.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52 (1), 3-50.

Ingwersen, P. (1999). Cognitive information retrieval. *Annual Review of Information Science and Technology*, 34, 3-52.

Interactive Services (2000) [Online]. Available: http://www.nielsenmedia.com/interactive/ [July 19, 2000].

Internet Demographics and eCommerce Statistics (2000) [Online]. Available: http://www.commerce.net/research/stats/stats.html [July 19, 2000].

Jansen, B. J. & Pooch, U. (2001). A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science*, 52 (3), 235-246.

Jansen, B. J., Spink, A. & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing & Management*, 36 (2), 207-227.

Järvelin, K. (1986). On information, information technology and the development of society: an information science perspective. In P. Ingversen, L. Kajberg & A. Mark Pejtersen (eds.), *Information technology and information use: Towards a unified view of information and information technology*. London: Taylor Graham, 35-55.

Järvelin, K. & Wilson, T. D (2002). On conceptual models for information seeking and retrieval research. To be published in *Information Research*.

Kelly, G. A. (1963). *A theory of personality: the psychology of personal constructs*. New York: Norton.

Kuhlthau, C. C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42 (5), 361-371.

Kuhlthau, C. C. (1993). *Seeking meaning: a process approach to library and information services*. Norwood, N.J: Ablex.

Kuhlthau, C. C. (1999). The role of experience in the information search process of an early career information worker: perceptions of uncertainty, complexity, construction, and sources. *Journal of the American Society for Information Science*, 50 (5), 399-412.

Kuhlthau, C. C. & Tama, S. L. (2001). Information search process of lawyers: a call for 'just for me' information services. *Journal of Documentation*, 57 (1), 25-43

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed). Chicago: University of Chicago.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (eds.), *Criticism and the growth of knowledge: proceedings of the International colloquium in the philosophy of science, London, 1965, volume 4*. Cambridge: Cambridge University, 91-196.

Large, A., Beheshti, J & Moukdad, H. (1999). Information seeking on the Web: navigational skills of grade-six primary school students. In L. Woods (ed.), *ASIS '99 : proceedings of the 62nd ASIS Annual meeting*. Medford, N.J. : Information Today, 84-97.

Lawrence, S. & Giles, C. L. (1998). Searching the World Wide Web. *Science*, 280, April, 98-100.

Lazinger S. S., Bar-Ilan, J. & Peritz, B. C. (1997). Internet use by faculty members in various disciplines: a comparative case study. *Journal of the American Society for Information Science*, 48 (6), 508-518.

Leventhal, L. M, Teasley, B. M., Instone, K. & Farhat, J. (1994). Age-related differences in the use of hypertext : experiment and design guidelines. *Hypermedia*, 6 (1), 19-34.

Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

Marchionini, G., Dwiggins, S., Katz, A. & Lin, X (1993). Information seeking in full-text end-user-oriented search-systems: the roles of domain and search expertise. *Library & Information Science Research*, 15 (1), 35-69.

Marchionini, G. Lin, X. & Dwiggins, S. (1990). Effecs of search and subject expertise on information seeking in a hypertext environment. In D. Henderson (ed.), *ASIS '90 : proceedings of the 53rd ASIS Annual meeting*. Medford, N.J. : Learned Information, 129-141.

Newell, A. (1969). Heuristic programming: Ill-structured problems. In Aronofsky, J. (ed.) *Progress in Operations Research, III*. New York: John Wiley & Sons, 360-414.

Nielsen, J. (1995). *Multimedia and hypertext: the Internet and beyond*. Boston, Mass.: AP Professional.

Nielsen, J. (1998) [Online]. *Cost of user testing a website*. Available: http://www.useit.com/alertbox/980503.html [March 18, 2002].

Nielsen, J. (2000). *Designing Web usability*. Indianapolis, Ind.: New Riders Publishing

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Paisley, W. J. (1968). Information needs and uses. *Annual Review of Information Science and Technology*, 3, 1-30.

Pharo, N. (1999). Web information search strategies : a model for classifying Web interaction?. In T. Aparac, T. Saracevic, P. Ingwersen & P. Vakkari (eds.), *Digital libraries: interdisciplinary concepts, challenges and opportunities: proceedings of the Third International Conference on the Conceptions of the Library and Information Science (CoLIS 3)*. Zagreb: University of Zagreb, 207-218.

Pharo, N. (2000). The preferences of information services by students in library and information science. *Norsk tidsskrift for bibliotekforskning*, 5 (13), 107-125.

Qiu, L. (1993a). Analytical searching vs. browsing in hypertext information retrieval systems. *The Canadian Journal of Information and Library Science*, 18 (4), 1-13.

Qiu, L. (1993b). Markov models of search state patterns in a hypertext information retrieval system. *Journal of the American Society for Information Science*, 44 (7), 413-427.

QSR N5 (2001) [Online]. Available: http://www.scolari.co.uk/qsr/qsr_n5.htm [November 1, 2001].

Rada, R. & Murphy, C. (1992). Searching versus browsing in hypertext. *Hypermedia*, 4 (1), 1-30.

Rasmussen, J. (1990). A model for the design of computer integrated manufacturing systems: identification of information requirements of decision makers. *International Journal of Industrial Ergonomics*, 5, 5-16.

Reid, J. (1999). A new, task-oriented paradigm for information retrieval: implications for evaluation of information retrieval systems. In T. Aparac, T. Saracevic, P. Ingwersen & P. Vakkari (eds.), *Digital libraries: interdisciplinary concepts, challenges and opportunities: proceedings of the Third International Conference on the Conceptions of the Library and Information Science (CoLIS 3)*. Zagreb: University of Zagreb, 97-108.

Rich, R. F. (1983). Management and problem solving styles: an assessment of information system designs. In A. Debons & A. G. Larson (eds.), *Information science in action: system design: volume 1*, Boston, Mass.: Martinus Nijhoff, 240-264.

Richmond, A. (1996). Enticing online shoppers to buy: a human behavior study. *Computer Networks and ISDN Systems*, 28, 1469-1480.

Saracevic, T. (1996a). Modeling [sic.] interaction in information retrieval (IR): a review and proposal. In S. Hardin (Ed.), *ASIS '96 : proceedings of the 59th ASIS Annual Meeting*, 33, 3-9.

Saracevic, T. (1996b). Relevance reconsidered '96. In P. Ingwersen & N. O. Pors (eds.), *Information science: integration in perspective*. Copenhagen: Royal School of Library and Information Science, 201-218.

Schön, D. A. (1990). *Educating the reflective practitioner*. San Fransisco, Calif.: Jossey-Bass Publishers.

Shneiderman, B. (1997). Designing information-abundant Web sites: issues and recommendations. *International Journal of Human-Computer Studies*, 47 (1), 5-29.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer interaction* (3rd. ed). Reading, Mass.: Addison-Wesley.

Shneiderman, B., Brethauer, D, Plaisant, C. & Potter, R. (1989). Evaluating three museum installations of a hypertext system. *Journal of the American Society for Information Science*, 40 (3), 172-182.

Silverstein, C., Henzinger, M., Marais, H. & Moricz, M. (1999). Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum*, 33(1), 6-12.

Spink, A (1997). Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science*, 48 (5), 382-394.

Strauss, A. & J. Corbin (1990). *Basics of qualitative research: grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Suchman, L. A. (1987). *Plans and situated actions: the problem of human-machine communication*. Cambridge: Cambridge University Press.

Tauscher, L. & Greenberg, S (1997). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47 (1), 97-137.

Taylor, R. (1968). Question negotiation and information seeking in libraries, *College & Research Libraries*, 29 (May 1968), 178-194.

Tuominen, K. (2001). *Tiedon muodostus ja virtuaalikirjaston rakentaminen: konstruktionistinen analyysi* [Knowledge formation and design of virtual libraries: a constructionistic analysis] . Espoo: CSC – Tieteellinen laskenta [Doctoral Dissertation].

Turoff, M. (1997). Virtuality. *Communications of the ACM*, 40 (9), 38-43.

Vakkari, P (2000). Cognition and changes of search terms and tactics during task performance: a longitudinal study. *Proceedings of the RIAO'2000 Conference*. Paris: C.I.D., 894-907.

Vakkari, P (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 7 (1), 44-60.

Vakkari, P. (2003). Task-based information searching. To be published in *Annual Review of Information Science and Technology*, 37.

Vakkari, P. & Hakala, N. Changes in relevance criteria and problem stages in task performance. *Journal of Documentation*, 56 (5), 540-562.

Wang, P, Hawk, W. B. & Tenopir, C. (2000). Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information Processing & Management*, 36(2), 229-251.

Wang, P. & Tenopir, C. (1998). An exploratory study of users' interaction with World Wide Web resources: information skills, cognitive styles, affective states, and searching behaviors. In M. E. Williams (ed.) *Proceedings of the 19th National Online Meeting*, Medford, NJ: Information Today, 445-454.

Wersig, G. (1971). *Information - Kommunikation - Dokumentation: ein Beitrag zur Orientierung der Infromations- und Dokumentationswissenschaften*. München-Pullach: Verlag Dokumentation.

Wersig, G. & Windel, G. (1985). Information science needs a theory of 'information actions'. *Social Science Information Studies*. 5, 11-23

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, 37 (1), 3-15.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55 (3), 249-270.

Wilson, T. D. & C. Walsh (1996). *Information behaviour: an interdisciplinary perspective*. University of Sheffield, Sheffield.

Yang, S. C. (1997). Information seeking as problem-solving using a qualitative approach to uncover the novice learners' information-seeking processes in a Perseus hypertext system. *Library & Information Science Research*, 19 (1), 71-92.

Appendix 1: Questionnaire

Students' use of the World Wide Web as a tool for information searching

This questionnaire forms one part of a larger survey of how the World Wide Web is used as a tool for working with comprehensive work tasks.

The survey consists of two parts, and I hope that *all* students will fill out the enclosed questionnaire and return it in my shelf on the fourth floor, preferably within the 20th of October.

For the second part of the survey we are looking for candidates that plan to use resources found via the World Wide Web as sources of information for their theses. If you belong to this group I would like to attend some sessions where you use the Web as tool in your thesis work. On these occasions I would like to perform an interview with you and, if applicable, other students that will take part in the session. I would also like to record the search session on video. Please note that it is the screen activity that will be recorded, not the students. I would, however, instruct you to talk aloud during the sessions, and to record your talking on the videotape.

If you plan to use the Web as part of your thesis work I would like to make an appointment with you. If you co-operate with other students on your thesis then I would prefer to make appointments with all of you. Unfortunately we cannot offer any financial compensation for participating in the survey, but among the participants taking part in the second part of the survey, three prizes consisting of *book cheques* will be drawn.

I hope that you will take the time to answer the questionnaire. Thank you very much!

Yours sincerely

Nils Pharo

Name:……. (please fill out for administrative reasons)
Class:…….

☐ Yes I will use the Web in my thesis work and would like to participate in the user survey.

I co-operate with:…….. (name/class)

Personal information
Gender (M/F):
Age:

Background
Previous education after high school:
_____
_____
_____
_____

Computer Experience
Years of computer experience: ____

On the scale below I would like you to make a mark to represent your experience with computers:

1_____2_____3_____4_____5_____6_____7
low high

For what purposes do you use computers:
_____
_____
_____
_____

Do you have a computer at home? (Y/N): ____

Internet experience
Years of Internet experience: ____

On the scale below I would like you to make a mark to represent your experience with the Internet:

1_____2_____3_____4_____5_____6_____7
low high

What Internet services do you use:
Electronic mail:____
Electronic mailing lists:____
World Wide Web:____
Newsgroups (Usenet News) :____
FTP:____
IRC:____

Other: :
_____

_____
_____
_____

For what purposes do you use the Internet:

_____
_____
_____
_____
_____

Do you have access to the Internet at home? (Y/N): ____

About your thesis

Have you decided what topic to write about? (Y/N): ____
If applicable, give a short description of the topic of your thesis:
_____
_____
_____
If the thesis will be written in co-operation, how many students will you co-operate with: ____

If you have decided what topic to write about, please mark the appropriate alternative below:

How did you decide the topic:
My own idea: ____
Proposal made by lecturer: ____
External task: ____
Other: _____

Previous research has shown that work with larger tasks, such as the writing of a thesis, often can be categorised into several stages. The following question is asked in order to identify the stage that you are in.

At what stage in the thesis work are you currently? (mark the best alternative, only *one* mark)

☐ I define the thesis

☐ I redefine the thesis

☐ I collect general information about the thesis' topic

☐ I explore and specify problems for the thesis

☐ I collect information for specific problems in the thesis

☐ I formulate the focus for the thesis

☐ I search for information to support the focus of the thesis

☐ I am at the stage of finishing my thesis

☐ I search for information in order to verify what I have done

Use of information sources for the thesis
I am interested in learning what kind of information sources you have used and plan to use when working with your thesis.

Below a number of alternatives are presented. I would like you to set a score for how you evaluate the individual information sources' importance *until now.* Please give a score in the range of 1-7 (where 1=not important, 7=very important)

Persons as sources
Lecturers at the department: ____
Fellow students: ____
Others at the department: ____
Friends: ____
External persons: ____
Others (please specify):
_____
_____

Library and library services
Loan of books: ____
Loan of article copies:____
Bibliographic databases: ____
Full text databases: ____
Citation databases: ____
Others (please specify):
_____
_____
_____

Printed material
Books: ____
Journals: ____
Newspapers: ____
Others (please specify): _____
_____

Internet services
Electronic mailing lists: ____
Electronic newsgroups (Usenet news): ____
World Wide Web: ____
FTP:____
IRC:____
Others (please specify):
_____

Other sources
_____
_____
_____

Use of information sources for the thesis (cont.)
Below a number of alternatives are presented. I would like you to set a score for how you evaluate the individual information sources' importance in what remains of your thesis work. Please give a score in the range of 1-7 (where 1=not important, 7=very important)

Persons as sources
Lecturers at the department: ____
Fellow students: ____
Others at the department: ____
Friends: ____
External persons: ____
Others (please specify):
_____
_____

Library and library services
Loan of books: ____
Loan of article copies:____
Bibliographic databases: ____
Full text databases: ____
Citation databases: ____
Others (please specify):
_____
_____

Printed material
Books: ____
Journals: ____
Newspapers: ____
Others (please specify):
_____
_____

Internet services
Electronic mailing lists: ____
Electronic newsgroups (Usenet news): ____
World Wide Web: ____
FTP:____
IRC:____
Others (please specify): _____
_____

Other sources
_____
_____
_____
_____

Appendix 2: Interview guide

Pre-search session questions/topics of discussion

1) Concerning the theme of your project, has any changes been made with respect to the goal or content? If so, what kind of changes has been made?
2) What have you been doing with respect to project work since the last time we met?
3) How do you feel about the progress of your project?
4) What sources of information have you been using since the last time we met?
5) Why do you want to use the World Wide Web today?
6) About today's search strategy: Are you looking for anything particular today or are you only looking (browsing)?
7) What is you goal of today's search session?

Post-search session questions/topics of discussion

1) How would you evaluate the results of today's session?
2) Are there other sources you could have used?
3) Are there other sources you planned to use before starting today's session?

Appendix 3: Classification rules for categories and attributes

1 Work task
The work tasks for all searchers were to complete their final year thesis at the LIS department at Oslo University College.

1.1 Work task goal/purpose
Information about the purpose or goal of the work task is collected from interviews with the searchers (actors). In addition, information from the questionnaires is used. During the search sessions the searchers may also tell about the purpose of their work task, and supplementary information may be collected from the final theses.

1.2 Work task complexity
We can classify the work task's complexity by identifying the workload demanded from the searchers. In the case of the LIS students they are supposed to deliver a thesis that can be characterised as having a small element of research and development (R&D) in it. This is the largest assignment given these students throughout their studies, and they are expected to deliver a thesis that has a certain amount of originality with respect to its content. In view of Byström's categories of task complexity the theses can be categorised as known-genuine decision tasks if the searchers follow the guidelines of the faculty. Thus only a small, if any, variance of complexity is to be expected.

1.3 Work task process stage
Kuhlthau's (1993) six stages of the information seeking process can be used to categorise the WT process stage. The search task goals and strategies of the searchers may indicate at what stage they are in the process, such knowledge can be collected from the interviews as well as the observational data. The searchers were asked in the interviews about how far they had come in the thesis work, all searchers indicated that they were in an early phase in their task performance.

2 Searcher (actor)
Both terms could be used to focus on the task as an act of performance, which do not only necessitate searching information. We have tried to use "searcher" throughout the chapter.

All searcher attributes can take positive or negative values. By this we mean that the searcher can have lack of task knowledge, search knowledge, search system knowledge, education, motivation, tenacity, uncertainty, and/or attention. This is, in particular, important to notice whenever one should state a relationship between a searcher attribute and another

attribute/category. A minus ("-") is used to indicate a lack in an attribute, e.g. "search knowledge -" is used to indicate that it is identified an instance of importance relating to the searcher's lack of search knowledge. Similarly the plus sign is used to denote a significant occurrence of presence of a searcher attribute, e.g. "motivation +" should be used to show that motivation is high in a certain situation or transition.

2.1 Work task knowledge

We know which out of two special subjects the searchers have chosen – information science or literature sociology. During the sessions they will reveal what they know, and what they do not know, about topics related to their task. Utterances related to such topics should be classified as task knowledge. In addition there are instances in the data that necessitates interpretation by the classifier, most important it will be to identify occurrences of "lack of task knowledge". This means that the classifier needs to possess a certain amount of task knowledge and be able to relate it to the searchers' performance. The "lack of task knowledge" is needed in order to understand why certain actions take place, or why the searcher has problems in understanding how to perform the task, it is unnecessary to record lacks in searchers' task knowledge when no consequences of this can be identified.

2.2 Search knowledge

(Lack of) search knowledge will be identified through the searchers' use of search engines and various forms of databases. In most cases this attribute is identified by looking at the actions performed by the searcher, but when the stories were written it was also made note of instances were the searcher clearly could have benefited from using search tools, but obviously did not – this was noted as an indication of lack of search knowledge.

2.3 Search system knowledge

Search system knowledge relates to the searcher's knowledge about the client program (in this case Netscape version 3.03). As with search knowledge this attribute is identified easiest by focusing on the searchers' actions, but the stories have also been written having in mind obvious lack of such knowledge – e.g. when the searcher does not know how to use the "edit-find"-option of the software and this could have been of help to her.

2.4 Education

All students are third year LIS students, but some also have other forms of higher education. We may find examples of how their education influence the search process, in particular indirectly, e.g. by influencing their task knowledge or search knowledge. Since their task is related to their education it may constantly play a role for the search process. For this reason it is recommended that its influence should be recorded whenever it can be identified as having a positive or negative effect on the interaction, if this effect is indirect the relationships should be

recorded, e.g. by pointing out that education affects (positively or negatively) the actors' search knowledge.

2.5 Motivation

During the process the searcher's motivation may change as result of different factors – this can be identified as utterances in various forms, e.g. as outcries or grunts. But motivational changes or effects can also be identified via its context, i.e. the searcher may decide to continue using a particular source or interaction technique longer than expected due to a heightened motivation. The opposite would be the case of a searcher losing her motivation and abandoning a situation earlier than one would expect. In the latter examples it is necessary to study the continuation of the session in order to identify whether the motivation has changed.

2.6 Tenacity

The tenacity may be hard to separate from the searcher's motivation, but we regard it as a more "stable" part of the searcher's personality. The tenacity is identifiable for searchers with extreme quantities of it, and is easiest identified in situations and transitions where the searcher invests a lot of, or little time and effort. The selection of particular ST strategies may indicate that the searcher has high tenacity, and searchers that are willing to spend much time waiting for slow servers also are considered to have this gift. We categorise tenacity as having effect only in such extreme cases, even though it may play a role throughout the entire session. The length of the session may also indicate the amount of tenacity held by the searcher, especially in cases were not much information is found.

2.7 Uncertainty (affection)

The searcher's emotions, which this attribute intends to cover, may dynamically change as the session progresses. We can identify such changes as utterances or actions performed by the searcher, in many cases the lack of actions – hesitation – is a sign of the searcher being uncertain about how to perform. Based on inspections of our data we have found it easier to identify "negative" than "positive" changes in a searcher's emotions, but "lack of" uncertainty should also be classified when it has an apparent effect on the interaction, or when it is possible to see that it is caused by specific characteristics of interaction. If a resource does not fit the searcher's mental model of web resources she may experience uncertainty because she does not understand how to interact with it.

2.8 Attention

A searcher's lack of attention can influence her ability to identify the content of a page or how to interact with a resource. It can also make her forget what she has recently learnt, or what she intended to do, on the positive side attentive searchers may uncover features of resources that many searchers would have problems to find. This attribute is easiest to spot at its

extremes, and should be recorded when it is apparent that the interaction in someway is hindered or supported by (lack of) searcher's attention.

3 Social/organisational environment

We focus on two attributes of the searcher's environment.

3.1 Actors

We use the attribute to record how people other than the active searcher(s) influence the search session. This may be fellow students, lecturers or persons that in some way are connected to the task and who have influenced the searcher's way of interacting with the web. We should code something as being influenced by "actors" when the searcher makes comments about this in the stories.

3.2 Goals and strategies

The goals and strategies of the environment within which the searcher performs her task indirectly influence the task, it is a theoretically derived attribute and should be coded whenever the searcher makes reference to it during the WIS processes.

4 Search task

4.1 Goal/purpose

The initial goal or purpose of the search task is revealed in the pre-session interview and recorded in the story analysis. During the session the goal/purpose may change - this should be recorded whether it is uttered by the searcher in response to results of her interaction or as a result of interpretation of her interaction over time, e.g. by examination of the query terms used or resource types inspected, compared to her initial goal.

4.2 Complexity/multidimensionality

The search task may contain a number of subtasks. This attribute is used to focus on that. Whenever it appears that the searcher switches from one subtask to another it can be categorised as a result of the ST complexity. Note that we do not use this category to classify switches between resources, e.g. from one newspaper to another. We have also chosen not to categorise the switch from searching for one topic to a very similar one, e.g. from one author to another.

4.3 Strategies

The search task strategy refers to how the searcher has planned to perform the search session. In addition to what we learn from the interview we look for patterns in the interaction in the form of conscious and systematic use of particular sources and interaction techniques. We can

also identify changes in the ST strategy as a result of the interaction, be that in the form of acquired information, "sudden" generation of new ideas, technical problems etc.

5 Search process

5.1 Transition

Search transitions are identified during the creation of the stories. Transitions are those parts of search sessions where the searcher looks for sources that she believes may contain information that can help her perform her task. Typically transitions take place in search tools and meta-resources, and searchers may switch between resources during a transition. Such a switch should however be considered starting a new transition in the following instances:

- When the searcher decides to start looking for another topic, e.g. if she has been searching for information about Ernest Hemingway and decides to search for John Steinbeck instead
- When the switch between search tools is the result of a conscious decision by the searcher, e.g. if the searcher decides to quit using Alta Vista in favour of Google to look for information on a topic. There are, however, examples of "unconscious" switches that should not be classified as generating a new transition, these are not part of the searcher's search task strategy. We give two examples here. If a searcher enters a query in a search engine and from the query results follows a link to a subcategory in a subject index or a similar resource, the use of this resource to look for relevant information should *not* be considered a new transition. Another example is the co-operation that exist between subject indices and search engines; the search engine Alta Vista and the subject index Yahoo used to co-operate so that if a query in Yahoo returned zero hits a new query would automatically be generated in Alta Vista and the results of this query be presented to the searcher. Today Yahoo co-operates in a similar way with the search engine Google. We believe it is fairly easy to identify whether such switches are made as a result of a conscious choice, if no such consciousness is identifiable then it should not be classified as a new transition.

The resource type is not a fundamental criterion for deciding whether some part of a session should be tagged a transition or a situation, the purpose of the searcher's interaction in a particular resource should be used to make that decision. If the searcher interacts with a search engine in order to study its functions, design or usability then those interactions are to be part of a situation. On the other hand, if she uses it to enter a query on a topic in order to find resources handling that topic, then it should be tagged a transition. Similarly, if a searcher uses a digital newspaper in order to look for a particular article or topic it is to be classified as a situation. If the searcher, however, uses the newspaper in order to find the link to the publishing house behind that paper, it is an example of a transition.

5.1.1 Action

We should classify any occurrence in the stories of the searcher following links, scanning pages, entering queries etc as “actions”. In the stories not all occurrences of actions have been recorded, in particular this should be noted in cases where a searcher traverses large hierarchies in a subject index like Yahoo. In such cases we have recorded that she traverses the resource, and this should be tagged with the action-attribute.

5.1.2 (Accumulated) results

The attribute is used to identify how results from previous interactions in the session influence the searcher’s current interaction. It is necessary to keep a session-oriented focus in order to identify such occurrences, not only should one look for traces of the interaction being influenced by previously acquired results, it is also important to “think forward” i.e. to keep in mind possibly future uses of the resources currently in use. Whenever use of previously acquired results is identified it should be tagged with this code.

5.1.3 (Accumulated) effort

The attribute refers to the investment of effort in individual transitions as well as the influence of the effort invested and accumulated throughout the session. The searcher may reveal signs on how her mental efforts during the session influence her current interaction. In many cases the searcher may feel exhausted from the interaction, and this will be uttered in the form of yawning, rubbing of hands, or similar “noises” that reveal tiredness. In such cases the attribute is easily identified and can be used to characterise the process. It is also advisable to look for signs of tiredness towards the end of sessions, at such stages it may influence the searcher’s attention and motivation, in the sense that the searcher makes mistakes during interaction or is less interested in exploring resources that might be of use to her. We should use “effort” to categorise our data whenever the searcher apparently invests much of it in a transition, which might be the case, to name a few, when her model of the resource (the so-called information space) is blurred, if her motivation is high, or if she decides to spend much time in a transition etc.

5.1.4 Information space

The searcher’s information space contains her model of the World Wide Web, including both *content* and *structure*. The model is not a mirror of the actual web, but is partly based on knowledge directly generated from previous Web use and partly based on presumptions deduced from her knowledge about the world outside the Web. This implies that both “accumulated results” and task knowledge influence her information space. We can to a certain point observe how the searcher’s information space is influenced by, and influence, her interaction, e.g. when she revisits web pages or when she interacts with unfamiliar resources.

We may also interpret the influence of the information space by observing effects on other attributes, e.g. changes in the searcher's search task strategy may be caused by a change in her information space, or a lack in the information space may cause a rise in the searcher's uncertainty. When categorising the information space we should be on the alert for the following phenomena, which both relate to the information space's model of Web *structure*:

- The searcher has problems in interacting with the resource due to its navigation facilities – indicates that the information space does not "cover" the resource;
- The searcher easily interacts with a resource that contains complex navigation facilities – indicates that the structure and navigation facilities are covered by the information space.

5.1.5 Time

The time factor is used to categorise particularly long transitions, but it may also be used in its "accumulated" form i.e. to identify whether interaction at later stages of the session is different than in the beginning. Note that differences in interaction may well be caused by a variety of attributes, such as ST strategy and remaining needs (see below), and that this attribute clearly is related to the "accumulated effort".

5.1.6 Remaining needs

The "remaining needs" can be identified by the searcher's comments. In the story analysis we focus on the remaining needs as a constraint that restricts the searcher's possibilities to use as much time as she wants and her abilities to explore as much of the Web as needed. The attribute is thus related to time and effort as well as searchers' ST strategies.

5.1.7 Resource type

There is a myriad of different resource types available on the Web. All transitions take place in one or more resource types and the resource type is decisive for the actions that can be performed with it – thus there is a strong relationship between these two attributes. For all transitions, and situations, each occurrence of a resource type, which may well be a part of a web site – e.g. a help page, or a page presenting the publisher, or a subject index's subcategory, should be tagged with this code.

5.1.8 Technical problems

This attribute is to be used to classify problems that occur during the interaction whether these are generated by the hardware, software, the specific client program, or the particular resource in use. Examples of such problems may be computer crash or software crash on the one hand, on the other hand it may be resources that contain erroneous code (e.g. HTML, cgi) or code causing problems for the searcher's interaction (e.g. redirect, frames).

5.2 Situation

Search situations constitute the parts of a search session during which the searcher explores resources in order to find information that she believes may help her execute or perform her task. A situation starts when the searcher accesses the resource and ends at her exit to another resource. The transition that connects the situations may be very short, e.g. the following of a link between to resources. On the other hand, a number of transitions, each of which may be consist of interaction with several resources, can take place between two situations. It is therefore necessary to be careful when identifying what is the purpose of the searcher's interaction with a given resource. Please look at the examination of the search transition above.

The attributes of search situations are similar to the ones that characterise transitions, with two exceptions; 5.2.6 relevance judgement and 5.2.7 relevance level. Below we go through the attributes. Whenever the attributes are similar we refer to the definition made under the respective transition-atribute.

5.2.1 Action

See 5.1.1 Action.

5.2.2 (Accumulated) results

See 5.1.2 (Accumulated) results

5.2.3 (Accumulated) effort

See 5.1.3 (Accumulated) effort

5.2.4 Information space

See 5.1.4 Information space

5.2.5 Time

See 5.1.5 Time

5.2.6 Relevance judgement

Whenever a searcher bookmarks a page during a situation we should classify it as a relevance judgement since it indicates that the searcher believes it may be of future use for her. Similarly it should be judged a positive relevance judgement whenever the searcher prints out a page, or makes note of what she learns from the page or says that it is of use to her. Negative relevance judgements are harder to identify as long as the searcher does not state that a page or resource is of no use to her. We advise that this attribute should be used to code positive relevance judgement and explicitly stated negative ones.

5.2.7 relevance level

The attribute is used to classify the level of relevance used by the searcher when she evaluated the web pages, as examined by Saracevic (1996) and Cosijn & Ingwersen (2000). We can use knowledge about the work task and search task to decide the relevance level. We believe that the default relevance level is "situational" (see below) whenever the searcher finds that a page is worth bookmarking. If a different level of relevance is identifiable then it should be made note of, e.g. if it is clear that the bookmarked page is to be included in a web site which is aimed at another group of users than the task performers it is clearly an example of "topical" relevance. If, on the other hand, the information in the bookmarked page is directly able to help the task performer in deciding how to structure the web site the relevance level is "situational". "Cognitive" relevance is the level in-between; the searcher is able to state whether or not a resource, a page, or something within a page is relevant to her task, but the "information" is not applicable to use in the solution of the task. These are the three levels of relevance that most probably are identifiable in the stories.

5.2.8 Remaining needs

See 5.1.6 Remaining needs

5.2.9 Resource type

See 5.1.7 Resource type

5.2.10 Technical problems

See 5.1.8 Technical problems

Appendix 4: Transcript of session no. 7.

The searcher's comments (in italics) have been translated for convenience with the original text in parentheses, while other comments made in this column have been kept in English. Text in brackets represent explanatory comments made by the researcher. Each appearance of a "." Is used to denote a pause of up to 5 second, i.e. "....." signifies a period lasting between 25 and 30 seconds without the searcher uttering a sound.

Action	**Time**	**Page**	**Comments**
		Department's main page	
Selects Internet search from embedded menu	0		…… [slow line]
Selects link		Excite - Netscape's internet search	
Scans page		Yahoo - Netscape's internet search	… [observer: please explain what you are doing] I though I would find the library here, let's see (Da tenkte jeg jeg skulle finne biblioteket her da, skal vi se)
Enters query: library	1.20	Yahoo - Netscape's internet search	Try to search (Prøve søk)
Scans page		Yahoo search results	[1233 hits in category, 8477 hits totally]. Let's se what kind of hits …… think I'll perhaps (Skal se hva slags treff …… tror kanskje jeg)
Selects back			Have to search for "public library" (må søke på "public library")
Enters query: public library	2.20		Let's see, got lots of hits on, let's see (Skal vi se, fikk så mange treff på, skal vi se)
Scans page		Yahoo search results	[38 hits in category, 1287 hits in total] . "internet public library"…
Selects link	2.51	Yahoo search results	
Scans page		Yahoo … San Francisco public library [1. Hit on list]	….. let's see, then I might, something saying "search only in San Francisco Public Library", so if I search for literature within this [category] (….. skal vi se, da kan jeg, noe som

			sier ”search only in San Francisco Public Library”, så hvis jeg søker på litteratur innafor det)
Enters query: literature [in category SF public library]	3.31	Yahoo … San Francisco public library	then I might be able to see what it is, gonna (så kan jeg se hva det er, kommer)
Looks at page		Yahoo search results	[0 hits] [did the candidate believe that the query was extended to other categories?] . no, that didn’t work (. nei, det gikk dårlig)
Selects back			
Scans page		Yahoo … San Francisco public library	……
Selects link		Yahoo … San Francisco public library	
Scans page	4.23	Yahoo: reference: libraries: public libraries	Try another library perhaps .. let’s see, yes (Prøve et annet bibliotek kanskje .. skal vi se, ja)
Enters query: literature [in category public libraries]	4.47	Yahoo: reference: libraries: public libraries	Try to limit by searching for literature within “public libraries” . (Prøve å begrense, med bare å søke på litteratur innenfor ”public libraries” .)
Looks at page		Yahoo search results	[0 hits] that was not particularly [good] (Det var heller ikke noe særlig)
Selects back	5.10		
Enters query: education [in category public libraries]	5.20	Yahoo: reference: libraries: public libraries	…
Scans page		Yahoo search results	[4 treff] …….. hm, [let’s] see (…….. hm, [skal vi] se…)
Selects link	6.28	Yahoo search results	..
Scans page		Rapides Parish libraries registering page	…
Selects link [visit without registering]		Rapides Parish libraries registering page	
Scans page		Rapides Parish libraries main page	…

Selects link		Rapides Parish libraries main page	Really have to check out this one first (Må nesten sjekke ut hva det her er for noe først)
Scans page		Page giving short introduction to the library with links to more on specific themes	….. hm ..
Selects link		Page giving short introduction to the library	
Scans page		Page presenting special collections at the library	…….
Selects link			
Scans page		RPL online – page with links to resources related to Louisiana	……. Hm …..
Selects link			
Scans page		Library lines – page presenting a regular column in local newspaper written by library staff	
Selects back			
Selects back			
			. Now I really should have added some bookmarks here [to be able] return (Nå burde jeg egentlig ha satt litt bokmerker her [for å] komme tilbake)[why does she wish to return here?]
Selects Internet search from embedded menu	10.15		
Selects link		Excite - Netscape's internet search	Let's see [observer asks her to perform in as natural way as possible]
Looks at page		Yahoo - Netscape's internet search	Let's see . (Skal vi se .)
Adds bookmark			
Looks at page		Yahoo - Netscape's internet search	.
Enters query: public library	11.11	Yahoo - Netscape's internet search	Let's see .. (Skal vi se ..)
Looks at page		Yahoo search results	
Adds bookmark			
Scans page		Yahoo search results	…..
Selects link	11.57	Yahoo search results	

		Yahoo … Internet public library	…. [observer asks her to talk loud] Yes, I tried to use "public library" once more, but, because I want to see if there are libraries that have made pages [similar to ours], and now I selected something called "Internet Public Library". I don't really know what that is, but I guess it's only a collection of things, but available on the Internet, no .. (Ja, jeg prøvde det med Public Library igjen da, men, siden jeg har lyst til å sjekke om det er noen bibliotek som har laget sider [slik som vi har?], og nå gikk jeg inn på noe som heter Internet Public Library, jeg vet ikke helt hva det er, men det er vel bare en samling med ting, men lagt på Internett, nei ..)
Selects link	13.11	Yahoo … Internet public library	Let's see, I'll try this one (Skal vi se, jeg prøver det her)
Scans page		Internet public library main page	… yes, "teen" "youth", I'll (… ja, "teen" "youth", jeg)
Selects link		Internet public library main page	Try "youth" and see where it brings me (Prøver meg på "Youth" og ser hvor jeg kommer)
Scans page		Internet public library – youth division's main page	.. "look around", let's see, there might be something here, help for school projects hm, some [icons] it looked a bit childish … (.. "look around", skal vi se, her kanskje det er noe, hjelp til skoleprosjekter hm, noen [ikoner] så litt barnslige ut …)

Selects link		Internet public library – youth division's main page	
Scans page		IPL youth division's reading zone, fiction and author resources	……..
Selects link		IPL youth division's reading zone	
Scans page		IPL youth division's reading zone – resources on mathematics	Doesn't appear to be much for the subject "Norwegian" here, but they seem to have something on mathematics … (Ser ikke ut som om de har noe særlig for norskfaget her, men det ser ut som de har noe med matematikk da …)
Selects back			
Selects back			
Selects back			
Selects link		Internet public library main page	
Scans page		Internet public library – teen division's main page	Here, there might be something, "books and writing", yes let's see . (Her kanskje det er noe, "books and writing", ja skal vi se .)
Selects link		Internet public library – teen division's main page	
Scans page		IPL teen division resources on Books and Writing	. Yes, this is really something that, it appears to be aimed atm older, than the page I looked at previously, so then, here are authors . (Ja, her var egentlig noe av det jeg hadde tenkt ja, så ut som det var for, eldre, enn den siden jeg var inne på i stad, så da, her er forfattere)
Selects link		IPL teen division resources on Books and Writing	Click on this one (Kikke på den)
Scans page		IPL teen division resources on authors	.. mhm .. yeas, is it possible to search, this was kind of interesting,

			put a bookmark on this one (.. mhm .. ja, går det an å søke, det var litt interessant, setter bokmerke på den her)
Selects back			I think (tror jeg)
Adds bookmark			[add bookmark on level above]
Selects link		IPL teen division resources on Books and Writing	
Scans page		IPL teen division resources on writing instructions	…. It's this way, I might return later to look at, and evaluate, how they have designed their pages (….Det er sånn at jeg kan gå inn seinere og kikke på, og vurdere, se hvordan, de har designet sidene sine)
Scans bookmarks			So, now we're really back where we searched for "public libraries", shall we (Så nå kan vi egentlig gå tilbake der vi søkte på "public libraries", skal vi)
Selects entry from bookmarks	17.54		See if we're able to find it, no! (Se om vi finner den, nei) [selects the page she has only recently bookmarked]
Scans bookmarks			Hm .
Selects entry from bookmarks	18.08		Here, I think (her, tenker jeg)
Looks at page		Yahoo search results [previously in the session bookmarked]	Let's see if I can find some other examples …… (Se om jeg finner noen andre eksempler ……)
Selects link	18.46	Yahoo search results	
Scans page		Yahoo …New York Public library	……..
Selects link	19.35	Yahoo …New York Public library	This one, I believe (Den, tenker jeg)
Looks at page		404 file not found	
Selects link [back to referring page]			
Scans page		Yahoo …New York	Hm …..

		Public library	
Selects link [several times]	20.15	Yahoo …New York Public library	Let's see …. This took some time … (Skal vi se …. Tok litt tid det her …)
Interrupts connection			No (Nei)
Scans page		Yahoo …New York Public library	.
Selects link	21.10	Yahoo …New York Public library	No (Nei)
Scans page		Yahoo … New York …Libraries	… [sighs] hm. [she has entered a local category for New York, *not* for libraries!]
Selects link	21.39	Yahoo … New York …Libraries	
Scans page		Yahoo main page	….
Enters query: college	22.11	Yahoo main page	Now I'll search for "college", let's see (Nå søker jeg på college, skal vi se)
Scans page		Yahoo search results	. It was a lot about sports …….. (Her var det mye sport …….)
Selects link	23.51		
Scans page		Yahoo … College and University Departments	…..
Selects link	24.15	Yahoo … College and University Departments	
Enters query: literature		Yahoo main page	
Scans page		Yahoo search results	[22 hits, no categories] …. [observer reminds her about talking aloud] I searcher for literature, I returned to the top of Yahoo, because I didn't think I found very much interesting when searching for, eh yes, "college" was what I searched for, perhaps I should try something different .. hm .. no (jeg søkte på litteratur, jeg gikk tilbake på toppen av Yahoo, for jeg syns ikke jeg fant så veldig mye

			interessant nå jeg søkte på, eh ja college var det jeg søkte på, kunne prøve noe annet kanskje .. hm .. nei)
Selects back	25.57		
Looks at page		Yahoo main page	Eh let's see (Eh skal vi se)
Enters query: authors		Yahoo main page	
		Yahoo search results	[43 categories, 2252 hits] At least I got lots of hits on literature, let's see . divided by genre . "young adults" …. Hm …. (Fikk i alle fall mye treff på om litteratur da, skal vi se . delt inn etter genre . "young adults" …. Hm ….)
Selects link	27.17	Yahoo search results	
Scans page		Yahoo … young adults : authors	.. No, it was a collection of authors, but it's not really targeted at, or well it might be targeted at youths, but not exactly the way I had thought . (.. nei, det var jo samling over forfattere, men den var vel ikke akkurat retta inn mot, eller den var vel retta inn mot ungdom, men ikke akkurat som jeg hadde tenkt det da .)
Selects link	28.03	Yahoo … young adults : authors	
Looks at page		Yahoo main page	I think I'll try "public library" once more, examine it a bit further (Tror jeg prøver public library igjen jeg, ser litt mer på den)
Scans bookmark	28.16		Look, this one I had bookmarked, hm (Se, den hadde jeg bokmerke på her, hm)
Selects entry from bookmark	28.21		
Scans page		Predefined Yahoo search on public library	…

Selects link	28.40	Predefined Yahoo search on public library	[selects entry no. 3 in the result list, the first two entries had been selected at a previous stage in the session]
Looks at page		Yahoo …St. Charles Public Library	.
Selects link	28.59	Yahoo …St. Charles Public Library	
Scans page		St. Charles Public Library main page	….
Selects link		St. Charles Public Library main page	
Scans page		Youth services - St. Charles Public Library	… look, there is something here (se her var det noe)
Selects link		Youth services - St. Charles Public Library	
Scans page		School services - St. Charles Public Library	….. [observer asks her to talk loud] no, it was, it said "help for youths", and then I thought it might be something on literature there, but there wasn't anything of particular, anything to click, so it was just information about what, probably what the library could assist on (nei, det var, det sto "hjelp til ungdom", så da tenkte jeg kanskje det var noe innen litteratur, men det var ikke noe å, særlig å klikke på videre her, så, det var egentlig bare informasjon om hva, sikkert hva biblioteket kan bistå med av)
Selects back			Help (Hjelp)
Scans page		Youth services - St. Charles Public Library	There is no such Internet help .. so it doesn't look like they've done much on (Ikke noe sånn internetthjelp .. så det ser ikke ut som de har gjort så veldig mye på)

Selects back			
Scans page		St. Charles Public Library main page	Internet issues, at least not here . let's see, "links to the Internet", that might be .. (Det med Internett, i hvert fall ikke der . skal vi se, "links to the Internet", det kan jo være ..)
Selects link		St. Charles Public Library main page	
Scans page		"Internet links for kids" - St. Charles Public Library	It was a bit childish, but . (Det var litt barnslig, men .)
Selects link		"internet links for kids"	Try books (Prøve books)
Scans page	32.16	Page with links to different kinds of resources for children	….. let's see, here, maybe this might be (….. Skal vi se, her kanskje [det] er noe)
Selects link	32.20	Page with links to different kinds of resources for children	Interesting, help in school work, hm (Interessant, hjelp til skolearbeid, hm)
Scans page		Kid's web – resource for children for help at school work	…..
Selects link		Kid's web	
Scans page		Page with links to resources within the Arts [music, drama, literature]	……
Selects link		Page with links to resources within the Arts	
Scans page		Page with resources within children's books	. hm . it looks like it's interesting, with links to, yes, books for kids, it's, it's perhaps a bit childish, but anyway, it's not particularly relevant for high school students but (. hm .det ser ut som det er noe interessant, med lenker til, ja bøker for barn, det er, det er jo kanskje litt barnslig, men hvertfall, ikke akkurat for videregående studenter men)
Adds bookmark			It's possible to take a

			look at it to see how they have (Det går an å kikke på den og se hvordan de har)
Scans page		Page with resources within children's books	Though about [designing] those pages, but it looks as if they have mostly links and no query facilities … (tenkt ut sidene her, men det ser ut som de har mest linker og ikke noen søkefunksjoner …)
Selects back			
Selects entry from bookmark	34.35		
Scans page		Predefined Yahoo search on public library	It seemed like I got the more interesting hits only by following the different links and look at what the libraries offered, so I don't really now where to go, but .. (Det virka som jeg fikk flest interessante treff med å bare gå inn på de forskjellige linkene og kikke hva bibliotekene hadde, så, vet ikke helt hvor jeg skal gå nå men
Selects link [no. 6]	35.21		[links 1-4 have been examined at a previous stage in the session] .
Looks at page		Yahoo … Saint Paul Public Library	..
Selects link	35.42		
Scans page		Saint Paul Public Library main page	……. [observer asks her to talk loud] no, now I've eh, went a, took one of those overviews over what they really offer, went down their page and looked at what it was, and it eh, yes and then I thought I'd try this one, it looks as if it might be links out on the Internet, made by the library (nei nå har jeg eh,

			gikk en, tok en sånn oversikt over hva de egentlig hadde der, gikk nedover siden deres og så hva det var for noe, og det eh, ja så tenkte jeg skulle prøve det her, det ser ut som det kan være linker ut til Internett da, som biblioteket har laget)
Selects link		Saint Paul Public Library main page	
Scans page		Saint Paul Public Library virtual branch [links to categories of resources]	Yes, let's see … .. "subjects" (Ja, skal vi se ….. "subjects")
Selects link		Saint Paul Public Library virtual branch	Take a look to see what it is, subjects (Kikke hva det er for noe, fag)
Scans page		Saint Paul Public Library subject page, links to pages with diverse subjects	. no, it wasn't that interesting really (. nei, det var ikke så interessant akkurat)
Selects back			
Selects back			
Scans page		Saint Paul Public Library main page	….
Selects link		Saint Paul Public Library main page	
Scans page		Page giving information about the redesigning of a room at the library	[scans quickly through page without looking at anything in particular, might be looking for links]
Selects entry from bookmark	39.08		
Scans page		Predefined Yahoo search on public library	.
Selects link	39.20	Predefined Yahoo search on public library	
Looks at page		Yahoo …DeKalb public library	.
Selects link	39.28	Yahoo …DeKalb public library	
Scans page		DeKalb public library main page	…
Selects link		DeKalb public library main page	
Looks at page		DeKalb public library – page containing links	[observer asks her to talk loud] Now I went to a

		within diverse themes [how to use the internet, various subjects etc.]	new library and then I found … something called links and then I though that it might have something to do with the Internet, let's (Nå gikk jeg inn på et nytt bibliotek og så fant jeg … noe som heter linker og så tenkte jeg kanskje at det hadde noe med Internett å gjøre, skal vi)
Selects local link		DeKalb public library – page containing links within diverse themes	see, here there are more links [the local link takes the searcher below the introductory text] (se, her er flere linker)
Scans page		DeKalb public library – page containing links within diverse themes	Oh, yes, it's just to read down … hm … (Åå, ja, det er bare å lese meg nedover …hm …)
Selects link		DeKalb public library – page containing links within diverse themes	
Looks at page		Links to resources that traditionally exist in printed form	..
Selects local link		Links to resources that traditionally exist in printed form	Literature (Litteratur)
Looks at page		Links to resources that traditionally exist in printed form – "literature" part	… Here there is at least something on literary resources on the Internet (… her er det i hvert fall noe med litterære ressurser på Internett)
Selects link	41.42	Links to resources that traditionally exist in printed form – "literature" part	Might see what it is, no [she thought she was following the wrong link] (Kan jo se hva det er, nei
Scans page		Literary resources on the Internet – main page	Yes ……. It's divided into, using, using (Jo ……. Det er delt inn etter, etter)
Selects link		Literary resources on the Internet – main page	periods, do it's (perioder, så det er)
		Literary resources on the Internet – Romantic	… but it doesn't really appear to be targeted at

			students, so then (… men det så ikke akkurat ut som det var retta inn mot elever, så da)
Selects back			
Looks at page		Literary resources on the Internet – main page	It gets a bit uninteresting .. (blir det litt uinteressant ..)
Selects back			
Session ends	43.30		

Appendix 5: Story representing search session 7

This is the first search session performed by this searcher.

About the work task

The student co-operated with three other students (*3.2 actors*) in creating a Web resource on fiction authors for pupils at the secondary education advanced level (*1.1 work task goal*). Pupils at this level are supposed to write a larger assignment and they very often choose to write about literary works and authors.). The task is proposed by the Oslo public library (*3.2 actor*), which is often contacted by pupils for such purposes (*3.3 organisational goal*). To decide which authors to collect data on the students have sent a questionnaire to various secondary level schools and discussed with representatives for the public library. They also visited a secondary level school to learn more about the topics of previous assignments, and in this way learnt to know a school librarian who helped them in deciding the kind of information that the pupils were looking for. Thus *3.2 actors* in a social environment within the LIS as well as secondary school *3.1 domains* with participants from different organisations function as *Wt resources* and influence the searchers. Searchers use these actors to obtain *2.1 WT knowledge*.

This is the first recorded session by this searcher, and she says that she feels that the project has not really started yet, one might say they are in the initiation stage of the project (*1.3 WT process stage*), that the students try to get a picture about how the task "ideally may be solved". Previous to the session the searcher has "searched a little bit" on topics related to the task.

About the searcher

The searcher claims to have slightly above average computer skills (4.4 on a 1-7 Likert scale), and average Internet skills (*2.2 search knowledge*). She has 7 years of computer skills and has used the Internet for two years, which indicates that she started using it when she attended the college. She uses email and Web regularly and the latter she says she uses to "search for interesting information" (*4.4 search system information*). She has Internet access in her own home. She has selected information science as her special subject in the third year, which means that she has more training in searching databases and query formulation than students without this specialisation have (*2.2 search knowledge)*. Her formal *2.1 WT knowledge* with respect to literature stems from her LIS education, which includes lecture on a variety of literature-related topics during the first two years of study.

With respect to information sources she thinks that her teachers, external persons (both groups being *3.2 actors in* her *social/org. environment*) (those giving the assignment) and the Web will be very important in the thesis work. Other sources of information she thinks will be

important are her fellow students, journal articles, and Usenet News. Books and newspapers will be somewhat less important sources (*1.4 task resources*).

About the search task

The searcher has decided to explore Yahoo to look for similar author indexes as the one they wish to make, i.e. indexes that contain entries to various aspects of the authors' authorship; interviews, book reviews etc (*4.1 ST goal → 5.2.4 information space*). She thinks that libraries and teachers at secondary level schools are the one most likely to have published such resources (*4.3 ST strategy*).

The session

The searcher starts the session by selecting "Internet search" (*5.1.1 action*) from Netscape's embedded menu (← + *2.3 search system knowledge*). The data transfer is very slow (*5.1.8 technical problems* - →) so it takes more than 30 seconds for the Netscape page to turn up (*5.1.5 time*). She selects the link to the Yahoo section in Netscape search (*5.1.1 action – 5.1.7 resource type*), this is where users can enter queries, which – when executed – bring them to Yahoo's ordinary result pages. The searcher explains that she wishes to look for libraries, to use this as a point of departure for author indexes (*4.3 ST strategy →*). She enters the query "library" (*5.1.1 action*), which returns 1233 category hits and 8477 real hits. She feels that this is too many hits (*5.1.2 accumulated results* - → *2.7 uncertainty/affection*) and states that "I will have to search for public libraries" and enter "public library" as her query term (*5.1.1 action – 5.1.7 resource type*). This returns only 30 category hits and 1287 "real" entries. She selects (*5.1.1 action*) the category entry on "San Francisco public library" and resonates that she can "'search only in San Francisco public library'". This is a misunderstanding since the query options is not in SF public library's web site but in the Yahoo category on the library (*5.1.4 information space* - →). That her mental picture of Yahoo and its search facilities is unclear shows her next *5.1.1 actions*; she selects back and enters the category Yahoo : … : public libraries and perform the same query (literature) in the category public libraries. She says that she wanted to narrow her search, but as this query is matched against the short descriptions of the individual libraries that fit this category it is much too narrow. She also combines this category with a query on "education", which returns 4 hits (*5.1.1 action – 5.1.7 resource type*). Transition 101_T1 ends and the first situation starts when she enters one of the links returned by the query (*5.1.1 action*). (6 min. 28 secs.)

The searcher would probably have benefited from selecting the category on authors and looked for indices on authors within this category (possible 5.1.1 action). She chooses as her point of departure a category (public libraries) which is probably too narrow for her causes, since categories that would have satisfied her need are probably categorised on their own. The searcher is obviously unaware of this (*2.3 search system knowledge* - → *5.1.4 information space* → *4.3 ST strategy).*

The resource entered in situation 7_S1 is made by "the Rapides Parish libraries" (*5.2.9 resource type*). She enters a page where one can make an optional registration of oneself, but chooses to continue without registering (*5.2.1 action*). She explores the resource's main page and selects a link that leads to a short introduction to the library (*5.2.1 action – 5.2.9 resource type*). She thoroughly (for almost 40 seconds) scans this page (*5.2.1 action – 5.2.9 resource type ← + 5.2.5 time ← + 5.2.4 effort ← + 2.5 motivation*) to see if it contains links to any resources similar to the one they are going to make (*4.3 ST strategy ← + 1.1 WT goal*) and selects a link to a page that presents the library's special collections, which is similarly explored, and from this a link to an index to Louisiana related resources (*5.2.1 action – 5.2.9 resource type*). Also this page is inspected with intense care (*5.2.3 effort + → 5.2.1 action*), and the searcher selects a link (*5.2.1 action*) to a page that contains a column regularly written by library staff in a local newspaper (*5.2.9 resource type*). The searcher comments that she perhaps should have added some bookmarks to the resource, but does not (unclear *4.1 ST goal -→* which influences *5.2.6 relevance judgement*). She ends the situation after 3 minutes and 47 seconds when she selects Internet search from the Netscape menu (*5.2.1 action*).

The searcher invests considerable 5.2.5 time and 5.2.3 effort on this resource, but there are no indications that it should contain any information relevant to the students' task (4.3 ST strategy). The fact that it was returned as a result of the searcher's Yahoo-query may suggest why she has such high hopes from it. The Yahoo query is however very general – a combination of public libraries and education should not be an uncommon theme. The searcher probably has problems in fitting the resource's content to her expectations (← 5.2.4 information space - → 5.2.1 action – 5.2.9 resource type) and this is why she explores it so thoroughly – to get an overview.

In Netscape's net search page (7_T2) she once again selects Yahoo as the active search tool and then bookmarks it (*5.1.1 action + → 5.1.5 time + 5.1.3 effort*). She then enters the query "public libraries" (*5.1.1 action – 5.1.7 resource type*) and bookmarks the search result (*5.1.2 accumulated results → 5.1.1 action + → 5.1.3 effort*). She selects the category Yahoo … Internet public library (IPL) and explains that "I don't know what this is, it's probably a collection of things put out on the Internet" (*5.1.4 information space -+ → 5.1.1 action*) and selects the links that leads to IPL's main page (*5.1.1 action*). Thus situation 7_S2 starts. (2 min. 56 secs.)

We see that the searcher rationalises by adding bookmarks to a result page that she probably wishes to return to later in the session.

In IPL the searcher scans the main page and decides to look at the "youth" section of the resource to which she finds s link (*5.2.1 action – 5.2.9 resource type*). She finds a link to "help with school projects", which she thinks might be useful and thus selects (*1.1. WT goal + → 5.1.1 action*). She finds links to resources on mathematics, but "nothing for the subject "Norwegian" [Sic.]" (*2.1 WT knowledge - → 5.2.4 information space*). She backtracks (*5.2.1*

action) to IPL's main page and selects the link to the "teens" section (*5.2.1 action – 5.2.9 resource type*), in which there is a link that she follows (*5.2.1 action*) to a section on "books and writing", which she finds interesting. It also contains a page directly related to authors (*5.2.9 resource type*), and she backtracks (*5.2.1 action*) and bookmarks the books and writing section (*5.2.6 relevance judgement*). She explains that she will return later to look at how they have designed their pages (*4.3 ST strategy ← + 5.2.2 accumulated results*), and now wishes to return to the query on public libraries, which she does with the help of the bookmark entry set in the previous transition (*5.2.2 accumulated results + → 5.2.1 action*). Situation 7_S2 lasts 4 minutes and 43 seconds.

In the teen section there were several other links that it might have been useful for the searcher to follow (2.8 attention - → 5.2.1 action) and possibly might have constituted a short cut for the page on authors. She does, however, bookmark a page, which may be used as a starting point for further exploration later.

She returns to the query results on public libraries in Yahoo (*5.1.7 resource type*) (7_T3) and via the Yahoo subcategory on New York public library (*5.1.7 resource type*) she selects the link (*5.1.1 action*) that apparently leads to its main page. The link does however lead to a "file not found" error, thus situation 7_S3 is classified as a *dead end*. (7_T3 lasts 1 min. 38 secs.)

The error message contains a link to the referring page (*5.2.9 resource type*) and therefore the searcher uses this (*5.2.1 action*) to return to Yahoo.

7_T4 takes place in the subcategory on NY public library which involves following a link that refer to the page it is embedded in (*5.1.1 action – 5.1.7 resource type ←→ 5.1.4 information space*)! It seems that the searcher is uncertain about how to use Yahoo's hierarchy (*2.3 search system knowledge - →*), which in turn influence her *5.1.4 information space - →* making her perform *5.1.1 actions* that returns little of interest. The searcher also selects a link (*5.1.1 action*) that leads to a subcategory for libraries in New York where New York is the super-category. She ends the transition by selecting a link in the Yahoo … New York … libraries-category (*5.1.1 action – 5.1.7 resource type*) that leads to Yahoo's main page. It is probable that part of her *4.3 ST strategy* is to look for large public libraries' web resources. (1 min. 57 secs.)

In transition 7_T5 the searcher enters the query "college" in Yahoo (*5.1.1 action – 5.1.7 resource type*), which generates "a lot of sport". She spends approximately 90 seconds scanning the result list (*2.6 tenacity + → 5.1.5 time → 5.1.1 action – 5.1.7 resource type*), but does not find any relevant entries and backtracks to the main page to perform a new query, which starts a new transition. (2 min. 36 secs.)

In the searcher's 5.1.4 information space it was probably not noted that colleges in the US are often very closely related to sport teams (← - 2.1 WT knowledge - →). The transition also

signifies that the 4.3 ST strategy is too weak given that the searcher ought to have come up with more specific query terms and combination of terms.

The searcher next (7_T6) performs the query "literature" (*5.1.1 action*) in Yahoo (*5.1.7 resource type*), but the result list is surprisingly short (22 hits and no categories) and she therefore enters "authors" (*5.1.1 action – 5.1.7 resource type*) instead. The latter query returns 43 categories and more than 2200 real hits. She selects the category "Yahoo … young adults : authors", but is disappointed to only find only "a list of authors … directed to youths" (*5.1.4 information space - → 5.1.7 resource type*). She says that this was not really the dimension she is looking for (youth authors) and therefore she (*4.3 ST strategy*) decides to re-try the publishing angle, meaning looking for public libraries that have produced something similar to what the students wish to do themselves (*← 1.1 WT goal*). (4 min. 1 secs.)

She uses the bookmark entry (*5.1.2 accumulated results*) to return to the query on "public libraries" in Yahoo (*5.1.7 resource type*) (7_T7) and selects a link (*5.1.1 action*) to St. Charles public library, via its Yahoo subcategory (*5.1.1 action – 5.1.7 resource type*).

St. Charles public library's web resource (*5.2.9 resource type*) is examined in situation 7_S4. In its main page there is a link to its youth services, which she selects (*5.2.1 action*). The page at first appears to be relevant, because she noticed that it offered some kind of help for youths, but as it had nothing to do directly with authors she devaluates it (*4.1 ST goal - → 5.2.6 relevance judgement*). She says that the library does not appear to offer much help on finding information on the web, but when she backtracks (*5.2.1 action*) to the resource's main page she find a link to an interesting looking page. The page turns out to contain "internet links for kids" (*5.2.9 resource type*) and although she claims it looks "childish" (*5.2.6 relevance judgements ← 5.2.7 relevance level*) she selects a sub-page on "books" (*5.2.1 action*) and from this page selects a link (*5.1.1 action*) to another resource (constituting Transition 7_T8 – 4 secs.) called "Kid's Web", which looks interesting because "it provides help in school homework".

It seems that the resource and its content (*5.2.9 resource type →*) changes the searcher's *5.2.4 information space + →* which in turn influence her *2.5 motivation + →*, and makes her reformulate her *4.3 ST strategy*. Situation 7_S4 lasts 3 minutes and 17 seconds.

The resources found in the library resource is not directly relevant, but it provides the searcher with a link that leads her to a resource that seem to satisfy her demand on finding something similar to the students own resource.

Kid's web is a subject index aimed at children (*5.2.9 resource type*) and the searcher traverses (*5.2.1 action*) it to a page that contains resources within children's books. Although the books are aimed at children the searcher says that it can be examined for design and structural purposes (*1.1 WT goal → 4.3 ST strategy*), she therefore adds a bookmark to the page (*5.2.6 relevance judgement*). Thereafter she ends the situation (7_S5) by selecting the query results

page bookmarked in transition 7_T2 (*5.1.2 accumulated results + → 5.2.1 action*). The situation lasts 2 minutes and 19 seconds.

In 7_T9 the searcher scans the Yahoo query results (*5.1.1 action – 5.1.7 resource type*) and selects the first entry that she has yet to visit (*5.1.1 action*), which is yet again a Yahoo subcategory (*5.1.7 resource type*), this time on St. Paul public library, from which she selects the link leading to the library's main page (*5.1.1 action – 5.1.7 resource type*). She tells about her coincident way of searching: "it seems that I got the most interesting hits by just following the different links and look at what the libraries had". The *← 4.3 ST strategy* was probably not well defined, but was a result of the searcher's *2.5 motivation* being influenced by the available *← + 5.1.7 resource types*. (51 secs.)

In situation 7_S6 the searcher first scans the library's main page in search for a relevant entry (*5.2.1 action – 5.2.9 resource type →*), she spends approximately 1 minute (*5.2.5 time*) before she finds a link that satisfies her expectations (*5.2.4 information space*). It turns out to be a subject catalogue, but the searcher does not find any relevant subject entries and says "it isn't particularly interesting" (*5.2.6 relevance judgement*) before backtracking to the resource's main page (*5.2.1 action*). She scans the main page (*5.2.1 action – 5.2.9 resource type*) once more before randomly selecting a link (*5.2.1 action*) leading to a page about the redesigning of a room at the library. She then returns to the Yahoo result list via a bookmark (*5.2.1 action*). The situation lasts 3 minutes and 26 seconds.

The selection of a random link can be seen as an act of frustration (5.2.3 accumulated effort affects - → 2.5 motivation). The searcher is unable to find the expected information in the expected sections of the resource and chooses a random link just to be sure that the needed information is not hidden there. In this case the fact that the term "design" was included in the link anchor probably made this link the most probable (5.2.1 action ← + 2.1 WT knowledge) of the remaining links. It can also be seen as a sign of the searcher having exhausted her 4.5 strategy and partly reached her 4.1 ST goal to continue the session.

In Yahoo's query results (transition 7_T10) the searcher selects (*5.1.1 action*) the category on DeKalb public library and then the link to the library itself (*5.1.1 action – 5.1.7 resource type*). (20 secs.)

The searcher explores the public library in a similar way as in the previous situation (*5.2.2 accumulated results + → 5.2.4 information space + → 5.2.1 action – 5.2.9 resource type*); she looks at their list of links to external sources and explains that she "found [a link in the main page] called links and thought it had something to do with [resources on] the Internet. She examines the link list and enters a sub-category on resources that traditionally exist in printed form (*5.2.1 action – 5.2.9 resource type*). In this list there are some entries that deal with "literary resources on the Internet", and the searcher selects a link to an external resource

on this particular topic. This link selection constitutes transition 7_T11 (4 secs.), which ends situation 7_S7 after 2 minutes and 18 seconds.

The resource she accesses in situation 7_S8 is an index to literary resources on the Internet sectioned into literary periods (*5.2.1 resource type*). The searcher seems to find it interesting, but claims that "it is not targeted at pupils" (*5.2.6 relevance judgement ← - 1.1 WT goal*), which is given as the reason for not bookmarking it.

It is a bit strange that this resource was not bookmarked since it might give the students some ideas about how to categorise a literary resource (5.2.7 relevance level). It may be that the searcher is bored with the topic (5.2.3 accumulated effort - → 2.5 motivation), or that she thinks that the two sources that she finds relevant during the session (in situations 7_S2 and 7_S5) are considered to be sufficient for their purposes (5.2.8 remaining needs → 5.2.7 relevance level → 5.2.6 relevance judgement). Resources that are less relevant do not meet the "relevance threshold" set by the resources already found.

She then decides to end the session, it having lasted 43 minutes and 30 seconds. She quitely leaves the room and closes the door.