Chapter 7: Explanation in Information Seeking and Retrieval

Pertti Vakkari and Kalervo Järvelin
Department of Information Studies, University of Tampere – FINLAND

1. Introduction

Information Retrieval (IR) is a research area both within Computer Science and Information Science. It has by and large two communities: Computer Science oriented experimental approach and user-oriented Information Science approach with a Social Science background. The communities hold a critical stance towards each other (e.g., Ingwersen, 1996), the latter suspecting the realism of the former, and the former suspecting the usefulness of the latter. Within Information Science the study of information seeking (IS) also has a Social Science background. There is a lot of research in each of these particular areas of information seeking and retrieval (IS&R). However, the three communities do not really communicate with each other. Why is this, and could the relationships be otherwise? Do the communities in fact belong together? Or perhaps each community is better off forgetting about the existence of the other two?

We feel that the relationships between the research areas have not been properly analyzed. One way to analyze the relationships is to examine what each research area is trying to find out: which phenomena are being explained and how. We believe that IS&R research would benefit from being analytic about its frameworks, models and theories, not just at the level of metatheories, but also much more concretely at the level of study designs.

Over the years there have been calls for more context in the study of IS&R. Work tasks as well as cultural activities/interests have been proposed as the proper context for information access. For example Wersig (1973) conceptualized information needs from the tasks perspective. He argued that in order to learn about information needs and seeking, one needs to take into account the whole active professional role of the individuals being investigated. Byström and Järvelin (1995) analysed IS processes in the light of tasks of varying complexity. Ingwersen (1996) discussed the role of tasks and their descriptions and problematic situations from a cognitive perspective on IR. Most recently, Vakkari (2003) reviewed task-based IR and Järvelin and Ingwersen (2004) proposed the extension of IS&R research toward the task context. Therefore there is much support to the task context, but how should it be applied in IS&R?

Aims and Focus. The present chapter therefore focuses on the following two questions:

1. What are the goals of IS&R research, in particular, what does the research want to discover/explain (as it appears on the basis of research done) and how could current research be enriched?
2. Do work tasks play a role in information access and how could their possible role be investigated? How could the task aspect enrich current research?

The goals of IS&R research may be classified as (a) theoretically understanding information seeking and retrieval in the form of models and theories, (b) empirically describing and explaining IS&R in various contexts, and (c) providing support in the design of information systems and information management in various contexts. The

whole area is rather pragmatic: many experimental (IR) studies aim at improving IR system effectiveness; many other studies are descriptive regarding the ways people access information (but these studies also have improving people's information access on their agendas).

Science seeks understanding of phenomena. Scientific knowledge grows, among other ways, by experimenting with observed and/or hypothesized relationships under varying experimental conditions, thereby arriving at confirmation, elaboration or refutation of the relationships. Explanation is a requisite for understanding. What kind of relationships are in focus in IS&R sub-areas? Are they closely related or vastly different, giving rise to either a unified Information Science or are they several disciplines with different agendas?

Approach. There is no distinguishing name for the methods employed in this chapter. This chapter first analyzes the frameworks of the three research areas for their research designs, especially the dependent and independent variables. It tries to explicate which kinds of hypotheses and theories are meaningful in each approach, how they relate to each other and how they could be enriched. In addition, this chapter examines work task based study designs giving two examples of IS&R studies which both have a work task context. The contributions of the chapter lie in (1) the analysis of the three areas, their enrichment and relationships, and (2) the examination of the role of work tasks in study designs. We do not suggest any specific studies to be completed nor provide any empirical results.

The Organization of the Chapter. In Section 2 we shall discuss an approach to the growth of knowledge in science. In Section 3 we analyze current research in IR, interactive IR and in Information Seeking, and discuss possibilities of enriching these efforts through the incorporation of new variables. In Section 4 we look at work tasks as a possible explanatory factor in IS&R. We consider three possible study designs where either the variation of recall and precision or the information access process is explained under varying conditions. We also discuss two sample studies in IS&R employing work tasks in different ways in their study designs. Finally, we discuss possible outcomes and their consequences of work task based IS&R studies. Summary discussion and conclusions follow in Sections 5 and 6.

2. An Approach to Growth of Knowledge

In order to be able to explore the research on information retrieval and seeking, and the possible relations between these fields, conceptual tools are needed. In the following we introduce Wagner and Berger's (1985) theory conception, and use it as a framework in analyzing what has been explained and by which factors in IRS. It is also used to explore the paradigmatic dimensions of various types of studies.

The growth of knowledge in science generally means the growth of theories (Balzer et al., 1987; Kuokkanen, 1992). In an applied field such as IS&R, system evaluation is a strong element in IR research. Thus, one could claim that better evaluation results, not growth of theories, is the criterion for growth of knowledge in IR research. However, it is evident, that one can achieve better evaluation results and better retrieval methods if one's tests are based on the accurate modelling of the phenomenon of interest. It is the underlying conceptualization of the study object with crucial

independent and dependent factors, which produces the evaluation results. Therefore, it is possible to apply the theory growth model for exploring growth of knowledge in IR research.

Theoretical growth can be assessed on three levels: on a metatheoretical level, on the level of substantial theories, and on the level of theoretical research programs (Wagner and Berger, 1985). Metatheories provide general theoretical perspectives on the broader field of study. They essentially contain ontological, epistemological and conceptual presuppositions of a very general nature. They provide a means of thinking about the research object, e.g. how information retrieval or information seeking should be conceptualized, i.e. what are their most important features. Thus, metatheories offer directives for actual theory construction including what are the most fruitful research questions and best methods for answering those questions. Laboratory model is an example of a metatheory or a paradigm for research in IR. For instance, excluding the user is a metatheoretical assumption guiding research it produces. Although Wagner and Berger (1985) claim that metatheories seldom grow due to their nature as *Weltanschauung (world view, which is time and place specific)*, they can be assessed by the growth of the theoretical research programs they generate (Vakkari, 1998).

Substantial theories deal directly with the concrete research object, e.g. information search process or indexing. They essentially contain a set of concepts and assertions relating the concepts in an account of some phenomenon in IR. They include hypotheses that can be tested and are intended as answers to particular research

questions (e.g., do weighted keywords retrieve more effectively than un-weighted ones?) The extent to which empirical testing provides support for the substantial theories determines the degree to which one may say that theoretical growth has occurred (Wagner and Berger, 1985).

Wagner and Berger (1985) call a family of interrelated substantial theories a theoretical research program. In addition to empirical support, the growth of substantial theories can be assessed by comparing theories of a theoretical research program. This is done by comparing their conceptual and factual similarity. The former means the similarity of the concepts in theories and the latter the similarity of the relations between concepts in theories. Thus, theories can grow by introducing a new concept or a new relation between concepts leading to a more precise or comprehensive theory. For example, in IR research introducing the concept of "term weighting" or "relevance feedback" has created growth. In information seeking the concept of "information types" has been shown to lead to growth in theories of task complexity and information seeking (Vakkari, 1998).

In sum, the growth of knowledge in research can be achieved by a larger empirical support, by introducing new concepts or refining old ones, or by introducing new relations between the concepts in a theory. A change in the structure of a theory (introducing a new concept or relation) typically creates the strongest conditions for the growth of knowledge (Kuokkanen, 1992).

In a study design, one may examine the interaction of several kinds of variables:

- *Dependent* variables – the variation of which is being explained.

- *Independent* variables – the ones systematically varied in order to see the responses in the dependent ones.
- *Controlled* variables – the ones fixed to prevent uncontrolled variation in the results. They are not as central in describing the research object as the two former ones.

The remaining variables stay as *hidden* variables in study designs and may sneak into research results by producing or hiding observable responses, especially when human actors are a part of the study object. IR experiments try to avoid this but relevance assessments are sensitive to them anyway.

To analyze the structure of propositions and theories in the field of IS&R, we focus on dependent and independent variables used in the studies. The variation of dependent variables is being explained by the variation of independent variables. These are the two main groups of factors, which represent the research object.

3. Current IR Research and Enriching Traditional Approaches to IS&R

3.1. The Lab IR Framework

The basic laboratory IR framework has no user involvement. This model suggests documents, search requests, their representation, the database, queries, and the matching of the two latter as foci of research and development. Methodologically, it also suggests relevance assessments, recall base construction and query result evaluation as foci of analysis (Figure 3.1)

What is being explained in the model? First and foremost, the variation of recall and precision under several experimental conditions mainly focusing on variations in document and request representations and their matching. Thus document and request representations and matching method are typical independent factors provided by the model. Admittedly, IR research has other tracks, such as question answering, topic detection and tracking, clustering, which do not precisely fit into the above model of document retrieval. However, similar analysis regarding scope and types of explanation applies to them as well. Contributions are mainly new techniques (i.e. variations in representation and/or matching) that, based on experimental evaluation, improve recall and precision under test conditions. It is hoped that these generalize to the many different user contexts.

Another track is methodological study in IR, which focuses on test collections (mainly scalability, as in TREC), evaluation metrics (e.g., discounted cumulated gain, Järvelin and Kekäläinen, 2002), or often somewhat critically on relevance assessments (i.e. types and levels). These contributions, however, seek to adjust the reliability and validity of the basic approach to explaining the variation of recall and precision.

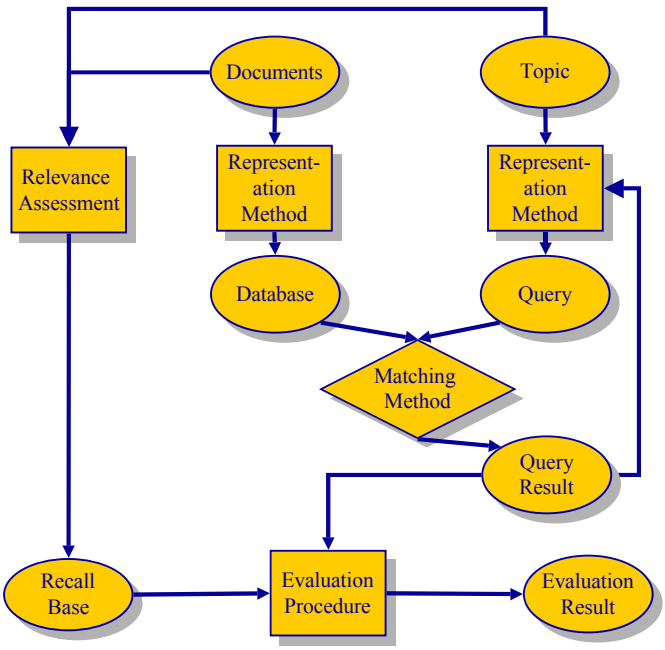

Figure 3.1. The laboratory framework of IR schematized

In the model (Fig. 3.1) the major independent factors (documents, topics and their representations) and also the central component of the dependent factor (relevance assessments) are taken as given. They are the sole source of variation of precision and recall. The variance of explaining factors is the natural point of departure for explaining the variance of precision and recall. However, it can be presumed that factors outside this model influence the variation of the independent factors, and consequently of the dependent factors. If it is possible to identify factors which cause systematic variation in independent factors and also in dependent factors, it is necessary to enrich the model by these external factors particularly if they help in enhancing the goal of IR evaluation, to improve retrieval effectiveness. This implies changes in the metatheoretical assumptions of the model. In the following we explore the possible external factors to the model, which may cause systematic variation in both independent and dependent variables.

Next we will analyze in more detail the major independent and dependent factors of the model and discuss the possibilities of enriching the model. After that we will apply a similar procedure to experimental studies on interactive IR.

3.1.1. Dependent variables

The phenomenon to be explained in this model is the variation of precision and recall in the retrieval of topically relevant documents. There has been much debate about the nature of relevance in information retrieval. Several scholars have argued that topicality is not a sufficient condition for a searcher to accept a document as relevant (Blair, 1990; Saracevic, 1976; Schamber, 1994). Usefulness may be one additional condition. The inquirer is looking for useful documents, the information content of which he can use in his task, which generated the search. Thus, the contribution of the information content to his task provided by a document is crucial for its relevance, i.e. for its utility. Therefore, a measure based on the utility of documents could be used as dependent variable. It could be measured as the number of useful items in the moment of retrieval or as the number of items finally used in the task for which information was pursued. Naturally, this would require new types of topics and utility assessments in the model.

It is difficult to say whether topical and useful documents to a task differ and whether this difference could be indexed. Are requests for topical and useful documents expressed differently so that we could infer new methods for representing requests? In general, would our ways of representing documents or requests and their matching change and in which way?

We doubt that it would be possible to distinguish topical and useful documents in terms of representation methods. It is difficult to suggest in which way they would differ from that perspective. Naturally, this is an empirical question, and remains open until closer empirical examinations. If it is not possible to distinguish these two types of documents for indexing purposes, then the current approach to assess documents in laboratory IR is sufficient.

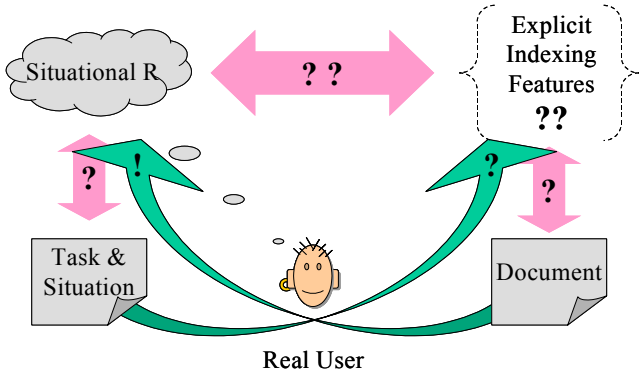

Figure 3.1.1. Situational R (relevance) in retrieval (Kekäläinen and Järvelin, 2002)

Some of the problems of using higher-order relevance in IR systems, or their design, may be discussed in terms of Figure 3.1.1 A real user, being thrown into a situation, may well be able to recognize a relevant document once presented (hence the exclamation mark). However, he may have difficulty in discussing the relevance criteria of the task and situation. Furthermore, he certainly has difficulty in expressing a request and formulating a query to the IR system, at least anything other than topical *as far as text is concerned* (except for bibliographic fields etc., if available, as discussed below), because current systems do not provide for anything else. The system designer probably never had the slightest idea of anything other than explicit topical indexing features, because there is no known pattern of situational indexing features that are explicit in text (the computer does not handle implicit features) and

useful to users (see Cooper, 1971). Therefore the available indexing features may not correlate to the situational relevance criteria, which the user did not express, save for one thing: topical relevance heavily correlates to situational as shown by Burgin (1992), and Vakkari and Hakala (2000).

In the current laboratory model the topical relevance of documents to a request is judged by external assessors by comparing the request and the content of the documents. Typically the threshold for accepting a document relevant to a topic is very low (Sormunen, 2002). This liberal relevance criterion is combined with a binary assessment scale. A document is either relevant or not.

The distribution of relevant documents in test collections is heavily biased towards the less relevant ones (Vorhees, 2000; Sormunen, 2002). Vakkari and Sormunen (2004) have shown in an interactive experiment that there was a significant difference in performance between two retrieval methods if all items assessed as relevant by external assessors were counted, but the difference disappeared if only those items were counted that were also identified as relevant by users. The users typically were not able to identify items of lower levels of relevance. Thus, there was a difference in effectiveness between the systems, but users were not able to observe it due to the marginally relevant documents in the result set. This suggests that topical relevance is defined too loosely allowing too large variation, and that it does not correspond to the way users conceive it. Therefore, it is difficult to say how validly the results obtained by loose relevance criterion are applicable to real life situations. It is suggested that this external factor should be incorporated in the laboratory model.

We suggest the extension of the dependent factor of the model into two directions. First, the utility of documents could be taken into account. Second, the variation of topical relevance should be decreased by tightening the relevance criterion. These solutions would increase the realism and validity of the laboratory model. This would create theoretical growth by introducing a new concept and refining an existing one in the model.

3.1.2. Independent Variables: Requests and Their Representations

In IR laboratory tests requests are typically expressed as topic descriptions for finding news articles. It is an open question whether requests from other domains for typical tasks or problem situations for other document genres differ from the paradigmatic requests of this dominating research approach (cf. Ingwersen, 1996). It is clear that document genres are different in these cases. If there is semantic or structural variation in expressions of typical requests between the domains, would it be possible to infer new and more effective methods of representing these requests or would current methods do? Before we answer these open questions, we cannot be sure whether our findings are also valid in domains other than those few used in our experiments.

Currently most of the topics are recall oriented. It is suggested that more emphasis should be given to precision oriented topics (Blair, 1990). Also other characteristics of topics like complexity seem to be related to retrieval effectiveness (Saracevic and Kantor, 1988), although this is not typically included in the laboratory model.

These reflections are supported by the findings in TREC experiments that topic variation affects much more than system variation on retrieval performance (Alemayehu, 2003). Retrieval effectiveness depends significantly more on which question is asked than on which retrieval system is used. Naturally, how the retrieval method deals with the question type affects performance. It seems that different retrieval mechanisms work relatively better on different topic types. This clearly shows that the way a request is expressed is a central factor affecting retrieval performance, and that some mechanisms are more effective in retrieving documents for certain types of topics.

These results suggest that a closer analysis and categorization of request types, and applying these results to request representation, would improve retrieval performance. Analyzing existing topics from the representation point of view would provide hints for designing particular retrieval methods for different types of topics. An option would be to categorize topics with regard to typical requests in certain domains.

3.1.3. Independent Variables: Documents and Their Representations

A presupposition of the laboratory model seems to be that all textual documents are similar so that all can be represented for IR similarly. It is expected that, e.g., variation between document genres (like news or scientific articles) or domains (like engineering or art) in vocabulary or in word distribution in documents, does not matter in how documents should be indexed. It seems to us that specifying documents and their representations by differentiating between documents of various genres and

domains is a fertile way of increasing the explanatory power of the model. This is an open question which can be answered by empirical studies. Evidently new IR techniques are needed, which react to the differences between genres.

Another more general presupposition is that keywords in documents are independent, although it is admitted that the assumption is not realistic (van Rijsbergen, 1980). However, most of the methods of representing documents lean on this assumption. It implies that keywords are basically treated as separate and atomic entities. Deviating from this assumption, a narrative text is a representation of a conceptually structured topic (Stubbs, 2001). Concepts and their relations construct the meaning of the text. Concepts are naturally expressed as words, and how these expressions are related to each other contribute to the meaning of the topic (Carter, 2000; Stubbs, 2001). Thus, the meaning of a word depends strongly on its relations to other words in the text.

If a document is not an arbitrary, but a structured set of words, should not documents also be represented accordingly for IR? Instead of being interested in the distribution of separate words, we would be interested in the distribution of a particular set of words representing a topic in documents or in document collections. There are some promising results produced by experiments designed from this perspective (Qiu and Frei, 1993; cf. Chung and Lee, 2004). Naturally, what is said about the representation of documents holds *mutatis mutandis* to requests.

A step further for representing documents as a set of associated words is to focus on the semantic role of particular words in expressing and elaborating the topic and respective sub-topics of a text. This can be called semantic-lexical representation of documents. Lexical cohesion arises from semantic relationships between the words (Stubbs 2001). All that it is required is that there should be some recognizable relation between them (Morris and Hirst, 1991). Lexical chaining is used for revealing lexical cohesion and semantically related words (Stokes, 2004). Lexical chaining has been applied successfully e.g. in text summarization (Barzilay and Elhadad, 1997), text segmentation (Stokes 2004) and improving precision performance in IR (Stairmand, 1997).

In addition to the approaches mentioned for the semantic-lexical representation of texts, discourse analysis and topical structure analysis (Georgakopoulou and Goutsos, 2001) may also provide conceptual tools for identifying different ways of developing a discourse topic, and analyzing how the words in this development are semantically related.

We propose the extension of the laboratory model to include various document genres in varied domains. It is also proposed that request representation methods other than the traditional bag of words approach, should be included in the model.

3.2. Experimental Interactive IR

Belkin (1993) proposed that the central process in information retrieval is user interaction with text, and that the user is the central component of the IR system. He ar-

gued that in traditional IR research little emphasis has been laid on developing representation schemes specific to information needs and problems. He further argued that one of the main goals of IR systems is to support inquirers in their information search process, especially in their interaction with texts. He stressed that understanding and representing users' information needs and problematic situations is important in developing tools for supporting the search process.

As suggested in the previous chapters the traditional laboratory model could be enriched by introducing representations of typical task and problem situations from various domains. This enrichment of the laboratory model does not include human actors and excludes studying the search process and various retrieval tools for supporting information searching. This can be studied only in interactive IR (IIR). Figure 3.2 depicts the elements of the experimental IIR model.

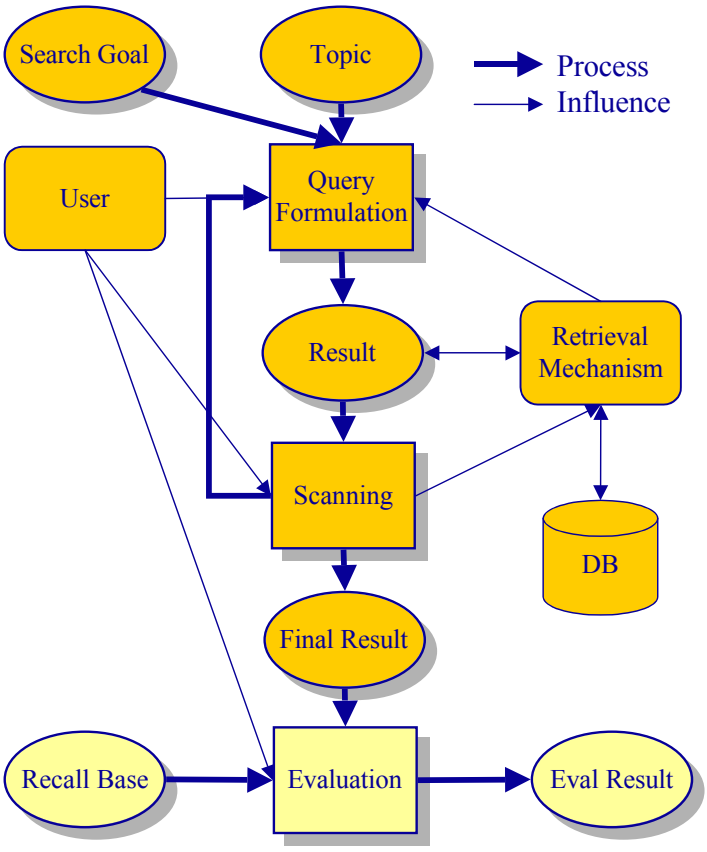

Figure 3.2 Schematized experimental IIR framework

3.2.1 Dependent Variables

In experimental IIR studies the dependent variables are precision and recall--the same as in laboratory studies. The recall base is formed in a similar way as in the laboratory model. Typically it is studied by determining how many predefined topically relevant items the searcher retrieves. Thus, the studies focus on users' ability to retrieve and identify topically relevant documents by using some retrieval mechanism. In some studies the satisfaction users experienced with search processes and outcome has also been studied.

One way to overcome the limitations of topical relevance is to study the impact of a retrieval mechanism. Hersh, Pentecost and Hickam (1996) suggest that information systems should be evaluated not only taking topical or situational relevance as the starting point, but in terms of system impact on the users. It is crucial how the information found by using a system helps subjects in their task for which they consulted the system. Hersh, Pentecost and Hickam propose outcomes-based methods to assess the impact of the system upon the user for particular tasks (e.g. for answering a question or solving a problem). Hersh, Pentecost and Hickam (1996) list several outcomes oriented studies in the field of medicine. The impact is typically assessed in laboratory conditions (cf. Section 4.3).

3.2.2. Independent Variables

Experimental IIR studies are typically designed for evaluating tools for supporting searching produced by laboratory IR in some user population (Beaulieu, Robertson and Rasmussen, 1996). The variation in the application of these tools has been used as an independent variable. Because the focus is on the influence of the tools on retrieval effectiveness, other factors in the research design have been controlled. Therefore factors other than tool related variables are rarely used as independent variables.

As in laboratory experiments it has been rare in IIR experiments to categorize search topics. Few studies include prior categorizations (Sihvonen and Vakkari, 2004; cf. Saracevic and Kantor, 1988) or *ex post facto* categorization for explaining the variation in results (Brajnik et al., 1996; Fowkes and Beaulieu, 2000).

Search goals given to the user have varied to some extent in experiments. Instead of finding documents on a particular topic, searchers are asked to identify different instances of an event type (Over, 2001). Goals other than recall orientation have rarely been used.

Provision of term suggestion devices based either on search results or knowledge structures have been the most typical independent factor in IIR studies. Provision of semiautomatic or interactive term devices based on relevance feedback on the one hand (e.g. Fowkes and Beaulieu, 2000; Koenemann and Belkin, 1996), and thesauri or ontologies on the other hand (e.g. Jones et al., 1995; Suomela and Kekäläinen, 2005; Sihvonen and Vakkari, 2004) or both (Joho et al., 2004) have been studied.

Also variation in query composition has been used as an independent factor. Supply of tools to support scanning and evaluating search results have been applied also to some extent as independent variables (e.g. Dumais et al., 2000).

Although user interaction with texts and system has been the focus of research, this interaction has rarely been treated as a process. How the independent factors influence search as a process has not been studied; typically only averages over the whole process have been given. For example, the search tactics used have rarely been studied in experiments with the exception of Ruthven, Lalmas and van Rijsbergen (2003) who incorporated this information into relevance feedback.

In all, it seems that various term suggestion devices have attracted most interest as independent variables in IIR experiments. For broadening the scope of IIR experiments, most arguments presented concerning independent variables in laboratory IR (sections 3.1.2., 3.1.3.) are also valid here. For example, it is evident that variations in the characteristics of topics, or in document genres and domains, or in search goals, influence how users search and utilize search support tools affecting retrieval effectiveness. Thus, extending IIR model by these factors would be beneficial. Section 4 provides additional argumentation and examples for expanding the range of independent factors in IIR.

3.3. Field Studies on IIR

The amount and range of field studies is so great that we give only a rough overview here. Field studies typically do not apply experimental research design. They use

user generated search topics and goals. The research design includes typically dependent, independent and controlled variables. The range of factors used as dependent variables varies from search process variables to search outcome variables like the number of relevant items. The range of independent factors is still more comprehensive varying from searchers' characteristics to search process variables. Controlled variables can be anything from searchers' characteristics to process variables. The one and the same variable may function in different roles in research designs depending on the research question. Borgman, Hirsh and Hiller (1996) discuss these different variables.

In field studies search effectiveness, i.e., the number of relevant items (or other measures inferred from them) are only one possible dependent factor. Also users' satisfaction with the search process or results has been used as dependent variable. It seems that search process variables are more common than outcome variables as dependent variables.

Searchers' characteristics, be they cognitive features or subject knowledge, have been popular as independent variables. Search task or goal, or especially the task that generated searching, has rather seldom been used as an independent variable. Search process variables like the choice of terms or tactics have gained more interest as independent variables. It has also been quite rare in field studies to include system features as independent or better intermediating variables in research designs.

In terms of variable combinations it has not been typical to include process variables as independent and outcome variables as dependent variables in designs. Studies, which combine searcher characteristics with process and outcome variables, are rarer

If the goal of information systems is to help actors in retrieving relevant items for some task and support in the search process, it is evident that both process and outcome variables are needed as dependent factors. However, because the outcome of searching is crucially affected by the search process and the task that generated searching, it would be fertile to use characteristics of tasks and their doers as independent variables and conceptualize process variables and system features as intermediating variables, which all influence search outcome. Naturally, it is a complex effort to include all these variables in a single research design. However, creating research programs for exploring and systematizing relations between these variable types would be a rewarding task. We discuss in more detail in section 4 the various variable groups by which to enrich research on information searching.

3.4. Information Seeking Studies and IR

In information seeking studies typical dependent variables have been the frequency of channel or source use and the preferences of channels in various groups. Thus, the main focus has been on the use and preferences of information sources and channels by actors. The use of information content obtained from those sources has been less intensively studied (Vakkari, 1997). Interest in the type of information searched for has gained some footing in recent years. Independent factors in these

studies have been mostly users' professional or demographic characteristics. Their information seeking has rarely been explained by the features of their tasks, interests or organizations (Vakkari, 1997).

By comparing the typical dependent variables in IR and IS studies it is easy to see that these two fields are interested in different phenomena. The former is focused on the precision and recall of retrieval methods used by humans in interactive IR. In field studies the search process has also been explored. Information content of items retrieved is included as representations of documents or requests in the form of keywords. The latter is focused on the preferences and use of channels. Information content has been touched in studies by categorizing information provided by the channels into broad subject classes.

The explaining factors rarely overlap in IR and IS studies. In interactive IR studies users' domain and search knowledge and their cognitive skills have been major independent variables, whereas in IS studies more general professional and demographic characteristics have been more common. Naturally there have some exceptions in both camps like studies by Ellis (1989), Byström and Järvelin (1995) or Wang and Soergel (1998).

If dependent and independent variables and their relationships do not overlap in the studies, it is understandable if representatives of both fields do not consider results and ideas from another field useful.

3.4.1. Kuhlthau's ISP Model: An Example of an Information Seeking Study

Next we will introduce Carol Kuhlthau's (1993; 2004) Information Search Process (ISP) model, which is one of the most influential models in the field of information seeking. We show how it has been extended to cover factors of interactive IR (Vakkari, 2001). The purpose is to illustrate that it is possible to utilize ideas and results of information seeking studies in studies on IR. The elements of the ISP model were used to explain how information searching in a database is related to task performance. Because our approach is based on task performance perspective, we also reflect on how the ISP model could be connected to certain characteristics of tasks.

What distinguishes Kuhlthau's model from most information seeking studies is that it is a process model, which connects source selection and the need and use of information to actors' changing understanding of their task. It conceptualizes information seeking as embedded in task performance and in information use. This is very rare in information seeking research.

The model explains what kind of information is needed at various stages of task process, how it is searched for and how it is used by the actors. The findings generated by the model are used for developing ways for mediators to intervene in the process for helping actors to cope with their information problems. They are also used for designing and developing library services to support users at their particular stages of ISP (Kuhlthau, 1993; 2004). It means that the findings have been used to build systems and mechanisms for supporting actors in finding useful information for their task performance. In that sense the goals of interactive IR and information seeking studies are not so far away from each other.

A central feature of the model is that it connects information searching closely to task performance, although it is not explicitly worded in this way. Information searching and use are conceptualized as integrative parts of task performance process. The model shows that phases in task performance differentiate the types of information needed and searched for and the major ways of searching. "Task" refers to non-trivial and mainly information intensive tasks.

The model consists of six stages. Each stage description consists of five elements: 1) the sub-task most appropriate to move the process on to the subsequent stage, typical 2) thoughts (cognitive), 3) feelings (affective), and 4) actions (physical), and 5) strategies to advance the process. As a whole the model is a description of the information search and use process for task completion. At each stage the sub-task leads to certain types of cognition and feelings, which in turn influences the actions taken and strategies chosen. Thus, over the process variation in some factors explains the variation of other factors implying that the whole process is strongly patterned. When actors' understanding of their task changes from vague to clear stage by stage, the type of information they are looking for and their actions to locate it and strategies to use it change also.

Vakkari (2001) has extended Kuhlthau's ISP model in the field of interactive IR (Figure 3.4.1). The independent variable in both studies was stages in task performance (ISP). Some of the dependent variables, like information types and search tactics used by Kuhlthau, were modified and refined for IR purposes. For example, in the

ISP model search tactics in information systems were categorized as browsing and querying, which is naturally too rough for observing in detail search tactics in databases. A more comprehensive and specific categorization of search tactics was developed. The ISP model was enriched by introducing variables describing the use of search terms and operators, and relevance assessments. The basic hypothesis for field studies was that the stages of task performance were connected to the types of information searched for, to the changes of search terms and tactics and to relevance judgements. These consequences from the ISP model were naturally based on the extension and refinement of its basic concepts for the purposes of IR.

KUHLTHAU'S MODEL	*VAKKARI'S HYPOTHESES*
Stages in ISP • Initiation • Selection • Exploration • Formulation • Collection • Presentation	*Stages in task performance* • Pre-focus - " - - " - • Formulation • Post-focus - " -
Types of information • General information (Background) • Specific information (Relevant) • Pertinent information (Focused)	*Types of information* • General information (Background) • Faceted background information • Specific information
Sources of information • Persons - Information systems	*Sources of information* • Persons - Information systems
Relevance judgements • Degree of usefulness	*Relevance judgements* • Degree of relevance • Relevance criteria used • Type of contributing information
Search tactics • Browsing or querying	*Search tactics* • A categorization containing 12 tactics
Search terms and operators	*Search terms and operators* • Number • Types (Synonym, NT,BT, RT) • Operator types
Mental models/Thoughts • General or vague - clearer or focused	*Mental models/Thoughts* • General or vague - clearer or focused

Figure 3.4.1 Main concepts in Kuhlthau (1993) and Vakkari (2001).

The empirical studies confirmed and elaborated the stated basic hypotheses (Vakkari, 2001; Pennanen and Vakkari, 2003). Their central findings concern actors' ability to structure their information needs and search topics, identify key words and formulate search tactics and assess the utility of the references retrieved at various stages of task performance. The more they advanced in the process, i.e. the clearer their understanding of the topic, the more able they were to perform the above-mentioned sub-tasks related to IR. The results generated some suggestions concerning systems design, although retrieval methods were not included in research design. This was conscious, because the authors first wished to understand IR as an ISP. The findings include ideas for developing and testing tools for supporting searchers with varying degrees of task knowledge.

It would also be possible to extend and refine the ISP model to cover more comprehensively some characteristics of tasks, e.g., task complexity and consequent information requirements as proposed by Byström and Järvelin (1995). As mentioned earlier the ISP model describes task performance from the angle of information searching and use. Also Kuhlthau (2004) has shown connections between the stages of ISP and task complexity. Therefore the suggested extension of the model would be justified.

We believe that this example shows that ideas and results from the studies on information seeking can be utilized in studies on IIR. It also shows that it is possible to enrich both research on information seeking and retrieval by conceptualizing them as part of task performance. This requires both theoretical work, which combines models, confirmed hypotheses and ideas from both fields, and research designs for validating the hypotheses derived.

4. Do Work Tasks Affect IS&R?

This section discusses explanations in IS&R from a task-performance point-of-view. While much of IS&R research is descriptive (how do people access information and what problems do they encounter?) and looks at IS&R through frameworks confined to IS&R, there are recent requests for a broader, work task-oriented approach (see, e.g., Vakkari, 2003). These raise significant questions for IS&R. Do work tasks affect IS&R in any significant way? Does information access affect task-performance? Which variables are there to be explained? Which possible explaining or intermediate variables are there? Does this really matter? If there are effects, how can we learn about them?

4.1. A Possible Framework for Analysis

Basically, we approach IS&R as embedded contexts of retrieval, seeking and work tasks (Figure 4.1). IR serves the goals of seeking, and information seeking the goals of the work task (or other interest). The same person symbol in all three contexts denotes the same or another actor(s) performing the work task, the seeking task and the retrieval task, interpreting the tasks, performing the process and interpreting the outcome, possibly resulting in task reformulation in each context. The person symbol in IR context signifies the possibility of applying human relevance feedback during a traditional two-run IR experiment as well as real longitudinal interactive IR over several short-term interactions. Figure 5 admittedly focuses on IR and neglects other means of information access. Furthermore, it is not analytic about significant factors affecting the seeking and work processes (see, e.g., Pharo, 2004; Vakkari, 2003).

However, possible evaluation criteria in each context are given at the arrows A – D. While many of them remain at quite an abstract level, they suggest types of variables, the variation of which one might want to explain. It allows for explaining variation in information access by organizational or work task criteria as well as explaining variation in work task performance by information access criteria. The following sections detail these possibilities.

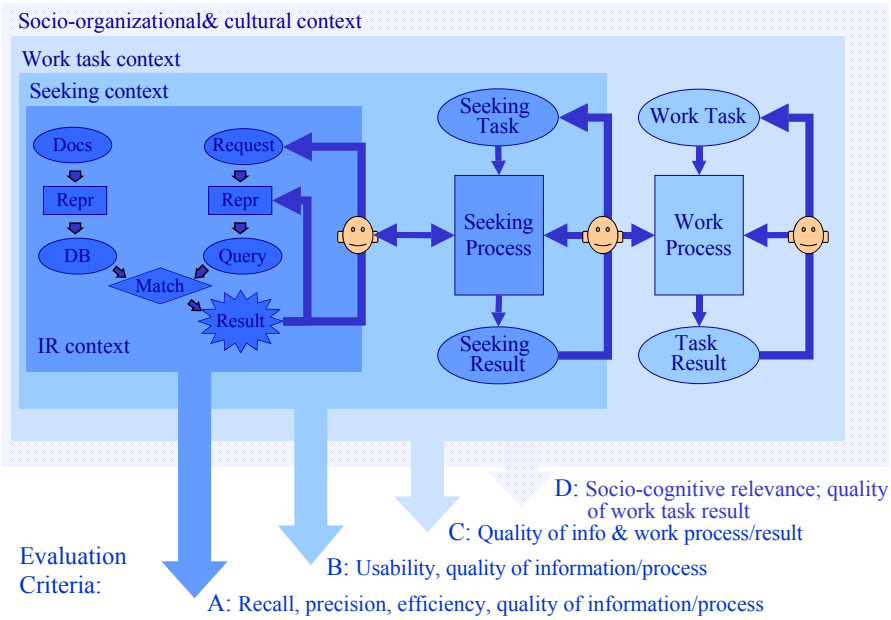

Figure 4.1. Nested contexts and evaluation criteria for task-based IS&R (extension of Kekäläinen and Järvelin, 2002; Järvelin and Ingwersen, 2004).

4.2. Some Possible Study Designs

Järvelin and Ingwersen (2004) propose nine broad dimensions that interact in IS&R processes. They are all found in the literature of IS&R but are hardly put together in any single study. In the following we present them briefly using a work task perspective.

1. The *work task dimension* covers the work task set by the organization, the social organization of work, collaboration between actors and the physical/system environment.
2. The *search task dimension* covers necessary seeking and retrieval practices, as understood collectively in the organizational practice.
3. The *actor dimension* covers the actor's declarative knowledge and procedural skills, and other personal traits as well – motivation, emotions.
4. The *perceived work task dimension* covers the actor's perception of the work task – forming the task that is really carried out.
5. The *perceived search task dimension* covers the actor's perception of the search task including information need types regarding the task and its performance process and perceived information space.
6. The *document dimension* covers document contents and genres and collections in various languages and media, which may contain information relevant to the task as perceived by the actor.
7. The *algorithmic search engine dimension* covers the representation of documents or information and information needs. It also covers tools and support for query formulation and methods for matching document and query representations.
8. The *algorithmic interface dimension* covers tools for visualization and presentation of information objects, collections and their organization.
9. The *access and interaction dimension* covers strategies of information access, interaction between the actor and the interface (both in social and in system contexts).

Each of the dimensions is complex and contains multiple variables. In any single study relevant variables need to be explicated depending on the goals of the study. It depends on the study design which variables are chosen as independent, controlled and dependent variables.

Based on the nine dimensions, we shall consider three different study designs, which incorporate work task effects. In the first study design we still remain within the experimental laboratory model of IR and seek to explain the variation of recall and precision under various experimental conditions, including realistic document collections, work tasks and relevance assessments. In the second study design we leave the explanation of the variation of recall and precision but seek to explain the access process under work task conditions; and in the third study design, we seek to explain the access process under actor conditions. The three designs extend the discussion of Ingwersen and Järvelin (2005).

4.2.1. Explaining the Variation of Recall and Precision Under Novel Conditions

This study design represents the basic laboratory IR evaluation design but uses test collections representing different parts of the real world. The standard IR test collections mainly contain news articles with little structural mark-up and few references; test requests that are very verbose, stable, conscious and topical; and relevance assessments that have hardly any work task scenario as their background (not representative, if any) and are very liberal (Sormunen, 2002). The evaluation scenarios require comprehensive (high-recall) retrieval.

The motivation of employing novel experimental conditions stems from the understanding that the variation of standard IR test designs grossly underestimates the variation in real-life IR settings. For comprehensive understanding the variations should match. The novel conditions therefore include the following:

- Real heterogeneous collections based on characterized work task contexts from organizations in varying domains.
- Real request sets based on characterized work tasks from the same organizations and subdivided into types (single answer / exhaustive; known item / factual / topical; specific / general; …).
- Relevance assessments by specialists in the tasks, however, retaining topicality but also considering other types of relevance (such as cognitive relevance assessed by task specialists (Cosijn and Ingwersen, 2001)).
- Evaluation reflecting work task (and request) requirements (number of documents required one / some / comprehensive, saturation, futility).

This remains under the old paradigm and is doable in the traditional way - even computer scientists can do it - but represents different parts of the reality with an *organizational existence warranty.* Therefore, one may observe in terms of the variation of recall and precision the dependent variables, whether earlier findings on the effectiveness of given techniques of document / request representation and matching hold for novel sub-domains of reality.

4.2.2. Explaining the Access Process Under Work Task Conditions

This study design represents a departure from the explanation of the variation of recall and precision. Instead, it seeks to explain how real people solve their information

problems related to different types of work tasks in realistic contexts. The underlying motivations for this effort include (1) the theoretical interest in better understanding the phenomenon of information access in work task contexts, and (2) the technical interest in developing novel and better kinds of task-based information access environments, which may require novel kinds of components. The study designs may be characterized by:

- The independent variables: a set of domain-specific work tasks analyzed for their features such as complexity (Byström, 1999), duration, types of difficulties (e.g., computational capability vs. open constraints), etc.
- The controlled variables: the organizational setting (e.g., a single large company); the information environment (e.g., all the shared information access tools the organization provides – some of which may be relevant for each work task).
- The dependent variables: the choice to use each access tool; the way each access tool is applied; the problems encountered; the kind of information sought with each tool (if any) (see e.g., Byström, 1999); the kind of information identified - utilized; etc.

The studies may be operationalized under two approaches: field studies and laboratory studies. In the *field study* version data collection happens in a real working organization. However, the type of organizations may vary between studies, e.g., governmental domain (e.g., law, banking) and information environments (e.g., comprehensive in-house information systems available). Likewise, the actors may vary in their expertise. The *laboratory study* versions of this approach distil findings from the field studies for more rigorous empirical testing. This may involve a collection of stan-

dard work tasks as routinely used in business schools (e.g., Isenberg, 1986) and an artificial work environment in the lab.

The contributions of this line of study are not of the standard IR type, which indicate how well a given type of system serves a standard type of de-contextualized retrieval task. Instead, they contribute an understanding of what kind of access tasks there are, how people try to solve them under the present ramifications, and what kind of problems they encounter and try to circumvent. Indirectly, the contributions cover issues like *'What types of systems might be useful for actors occupied by given types of work tasks?'*

4.2.3. Explaining the Access Process Under Actor Conditions

This study design continues the departure from the explanation of the variation of recall and precision. This effort seeks to explain how different types of real people solve their information problems related to controlled types of work tasks in controlled realistic contexts. The underlying motivations for this effort are as above in Section 4.2.2. The study designs may be characterized by:

- The independent variables: the actors performing the work tasks divided into groups in relation to work task performance (e.g., experienced professionals vs. novices regarding the work tasks, or one group receiving information access training plus a control group).
- The controlled variables: the organizational setting (e.g., a single large company), the actors performing the work tasks (e.g., all experienced professionals regarding the work tasks); the information environment (e.g., all the shared

information access tools as above); a set of domain-specific work tasks controlled for their features – a representative mix or fixed type.

- The dependent variables: the choice to use each access tool; the application way of each access tool; the problems encountered; the kind of information sought with each tool; the kind of information identified and utilized; the quality of work task outcome.

Also these studies may be operationalized under two approaches, field studies and laboratory studies. In the *field study* version data collection happens in a real working organization, which may vary between studies. Likewise, the actor groups (division principle) may vary between studies. The *laboratory study* versions of this approach distil findings from the field studies for more rigorous empirical testing. This may involve a collection of standard work tasks as routinely used in business schools (e.g., Isenberg, 1986) and an artificial work environment in the lab with controlled access tools (see below Section 4.3.2).

The contributions of this line of study are as in the preceding section. This time the findings on information access variations are explained by actor characteristics, not task characteristics. Indirectly, the contributions cover issues like *'How different types of actors might best be augmented for performing given types of work tasks?'*

4.3. Two Sample Studies

In this section, we discuss two studies that can be seen to utilize a task-based approach to IS&R. Neither study is based on a conscious task-based approach to

IS&R, however, and one of them is from a different discipline. Yet both studies demonstrate what a task-based approach to IS&R may contribute to IS&R. Harri Laitinen's (1996) study is a M.Sc. thesis in Information Studies investigating the use of an electronic high school learning material collection when answering high school exam questions. D. Isenberg's study is an investigation of business problem solving and compares experts to novices in the problem solving process and its outcome.

4.3.1. IS&R for High School Exams

Laitinen (1996) obtained a collection of high school textbooks on a CD-ROM from a Finnish publisher (Otava). The CD-ROM was equipped with a Boolean search system and also supported browsing the hypertext structure of the textbooks. He set out to analyze how 2nd year high school students used the CD-ROM when answering exam questions. The research questions contained, among others, the following:

- Do process-related features (e.g. search problems) relate to task outcome (the quality of answers)?
- Does the wording of exam questions affect the search process?
- Do students with different computing skills also differ in the task outcome?

The study design contained national exam questions in Finnish high school subjects (religion, physics, biology, psychology, and history), for a total of 12 questions. There were two versions of questions, one set quite verbose hinting at suitable search keys, and the other considerably briefer. The students' task was, using the CD-ROM and search system, to answer one (or more) exam questions within a time allowance of one lesson (45 minutes). The answers were assessed by a teacher using a standard mark scale of 0 – 6 points (a zero meaning a bad answer). Thus, the assumption was

that the use of the electronic collection contributes to the assigned (but very realistic) work task outcome. The study subjects were 29 second year high school students (12 girls, 17 boys), among whom 26 produced notes on their searching and 23 produced answers to one or more exam questions. The system was introduced to the students prior to the test. Multiple data sets were collected: search logs, student interviews, students' search notes, exam question answers, the teacher's assessments and the CD-ROM. Students produced answers to 6 different questions.

The findings of Laitinen's study indicate, among others, that the students mainly used briefsearch based on the exam question wording, but with an intervening examination of articles. Query formulation seemed difficult with little expansion or use of operators. However, the problems encountered in the search process seemed not to be connected to the marks obtained. The average marks differed between the two versions of question sets, the harder ones got lower than average marks. In general, the boys were more experienced in PC use and got better marks.

Laitinen's study reports a comprehensive analysis but provided insufficient data for definitive conclusions. It would also have been interesting to compare the findings to another group of study subjects, *not* having access to an electronic collection but rather to the same content as a set of printed text books. Nevertheless, the limitations may be overcome in later studies with more resources. The precise findings or their reliability are not the focus of our argument but, instead, the types of research questions one may answer using an approach like this. Laitinen's approach is an example of task based evaluation of information access methods and can, in principle,

trace the access method effects all the way to work task performance. In a comparative study based on this approach, one may explain the work task effects of information access methods.

4.3.2. IS&R in Business Problem Solving

Isenberg (1986) studied information seeking and use by experts and novices in business problem solving. His goal was to find out what cognitive processes managers use when they deal with business problems, and about differences in problem solving, and in the quality of the outcome, between experts (experienced managers) and novices (business school students). The experts were 12 experienced general managers from six corporations and the novices were three college undergraduates. All solved the same business case. The case was presented in a controlled way on seven cards, arranged randomly to reflect real business situations where cases do not begin in orderly fashion. The subjects could utilize the cards as much as they wanted without any extra costs or penalties. Data collection and analysis was through thinking aloud and protocol analysis.

Isenberg found that experienced managers made less use of the information available than novices. Instead, they made more conditional conclusions, which were based on less information supporting them. Experienced managers reasoned from general to specific, based on their experience, and considered several alternatives simultaneously. Powerful deduction and interpretation was typical for them. Isenberg explained this through opportunistic reasoning: if information is valuable but scarce, and its availability unlikely, an expert has learned to distil everything possible out of

the information at hand. Therefore the managers made speculative but plausible inferences even on a narrow basis. In the course of problem solving the managers started outlining solutions at an earlier stage than novices and produced qualitatively better results. In other words, experienced managers found a focus (Kuhlthau, 1993) earlier, perhaps directly without any preceding stages, while novices explored each case longer by identifying major concepts and relations before finding a focus. Regarding task complexity, the managers' perceived their tasks as considerably simpler than the novices, with the process, outcome and required information definable at the outset, while the novices had to construct these.

In these findings we see that even easily available information may be neglected (rightfully) as one may draw on one's experience (cognitive structures) and learned interpretation of one's situation. The way an assigned work task is perceived and thus formed into a personal work task depends on the actor's knowledge, which also affects the need for any additional information. While the assigned work task for the two groups was the same, the constructed personal work tasks were different (due to different levels of expertise) and, consequently, also IS&R was quite different between the two groups.

4.4. Possible Outcomes: Do Task-based Experiments Matter?

From an organizational work task perspective IS&R is most often just a supportive function, and its costs in terms of money and effort should be justified by the quality of the deliverables for some necessary and important work tasks. Let us assume therefore that two information access methods (different in some interesting aspect,

e.g., collection, IR engine, or interface functionality) are compared from a work task performance perspective in a knowledge intensive task domain. Quality of processes and outcomes, operationalized somehow, are assessed. An interesting difference between methods may be identified either due to its arguable potential contribution to task performance (~ technology assessment) or due to the large research efforts it has received in the IS&R community (~ a test of ultimate relevance of research efforts). In the following we consider two possible (abstracted) outcomes from the comparison and their consequences.

Outcome I: The performance variables show no difference between the access methods. Assuming a valid study design one has to conclude that the two information access methods compared are not significantly different from the work task performance viewpoint. The differences between them are not relevant even if they may have (or may have not) yielded different output and information finally used.

Consequence I: There are three possibilities. 1) Access to information is not significant in task performance, i.e., the quality of the process and its outcome are mostly determined by other factors. In this case the development of information access methods for the task domain is not justified. 2) Alternatively the comparison was unsuccessful in its design: information access does indeed matter, but the comparison did not consist of relevant alternatives. In this case, if the IS&R community does want to understand what in information access contributes to work task performance, one must identify relevant differences between access methods for further testing. 3) It is possible that the information access methods could not have done a better job; the task performer receives excellent service from the methods for the task (as long as it

remains stable). Also in this unlikely case further development of information access methods for the task domain remains unjustified.

Outcome II: At least some of the performance variables show a significant difference between the access methods. Assuming a valid study design one has to conclude that the two information access methods compared are significantly different from the work task performance viewpoint. The differences between them are relevant even if they may have yielded the same output and information used.

Consequence II: The comparison was successful in its design: information access does indeed matter in task performance in the given domain and the comparison did consist of relevant, significantly different, alternatives. In this case, the IS&R community has identified a difference in information access that contributes to work task performance and may utilize it in theory building and, instrumentally, in building better access methods.

5. Discussion and Conclusion

We have explored the scope and structure of propositions and theories, which have been mostly implicit, in studies of information retrieval and information seeking. We have also analyzed metatheoretical assumptions in these theories.

We showed that in studies of IS&R both dependent and independent variables differ considerably. In IR, particularly in experimental studies precision and recall are the dominating dependent variables, whereas in IS the intensity of use of various chan-

nels is the most common dependent factor. Users are excluded in Lab IR, but in IS human characteristics are typically independent variables. The independent variables in field studies of IR also differ considerably from those used in IS. These two fields seem to be interested in totally different phenomena. It is very rare to find studies which integrate variables from both camps.

In general, it seems that there has been theoretical growth within the laboratory IR framework in the sense that new concepts like term weighting or relevance feedback explain a significant proportion of the variance of recall and precision. In addition to this conceptual growth several empirical tests have supported the explanatory power of new concepts.

In Lab IR, theoretical growth has happened within its traditional domain by refining the traditionally central concepts by new ones representing new variations of IR techniques (for representation or matching). There have been neither radically new concepts (nor the associated relationships) for some while. For example, introducing the concept of an independent human actor into the scene would require redrawing the framework and would severely question the generalizability of the findings. Such a radical change has been discussed with much reluctance in IR, in principle admitting its necessity but in practice changing very little.

Scientific theories are needed for the following functions (Bunge 1967): *(i)* systematization of knowledge, *(ii)* guiding research, and *(iii)* mapping a part of reality. Frameworks and theories of IS&R cannot guide research outside their domains. Since there

are relevant phenomena outside their domains, framework/theory enrichment becomes necessary. All enrichment does not nicely fit in the current frameworks but, instead, may require redrawing them.

The Lab IR framework includes metatheoretical directives, which heavily restrict the range of independent variables. The framework as such does not deny the study of the effects of external factors, but neither does it help in this. Although it can be debated (cf. Kekäläinen and Järvelin, 2002), we believe that the minor or non-existent role human actors and their situations in the framework considerably restrict the generalizability of the findings to the real world.

We believe that enriching the framework by introducing new concepts representing actors, their typical domains and tasks would enhance the explanatory power of the framework. For instance, the range of topics and search goals could be broadened to reflect the scope of typical human tasks and respective information needs more validly. This is supported by the finding that topic variation explains retrieval effectiveness five times more than system variation (Alemayehu, 2003). The framework could also include a more comprehensive range of genres and domains of documents and their representation methods.

There has been much debate about the nature of relevance in IR, and consequently about the dependent variables in IR. In lab experiments recall and precision are exclusively based on topical relevance and not the utility of documents. It is difficult to

say whether retrieval by document utility rather than topicality could in any way be supported in document indexing. It is an open question.

In interactive experimental studies precision and recall have also been typical dependent variables. It has been proposed (Hersh et al. 1996) that the impact of the systems on the supported task should be taken into account. This would be a step forward for enriching the framework for interactive IR.

In IIR experiments, the application of retrieval mechanisms has been the most common independent variable. What has been said about enriching the laboratory framework is also valid here. Variation in the characteristics of topics, search goals and document genres would be beneficial. It is known that these factors influence how users search and utilize information (Blair 1990).

The range of dependent, independent and intermediating variables vary considerably in the field studies of IR. There are three typical limitations in these studies. First, typically they do not use system features as independent or intermediating variables. Therefore it is difficult to conclude how information search behaviour is related to the systems which are built to support it. Second, it is not common to include in the research designs user characteristics as independent variables, dimensions of search process as intermediating variables, and the features of outcome as dependent variables. Therefore it is complicated to predict how these various elements of searching are related. Third, task performance, and consequently the information searching it generates, is a process. This means that the search topic, goals and process gradu-

ally change when task performance proceeds (Vakkari, 2003). However, there are few studies which have focused on this process.

Our analysis shows that task and actor features may be incorporated in study designs either indirectly, i.e. through carefully selected test collections, or directly through carefully varied work tasks / test person characteristics. The former supports the continuation of the lab type of investigations, but with more varied test collections. The latter requires extended interactive study designs which consider actor, problem and task features. In both ways it is possible to improve the transferability of IS&R research findings to the real world as usable systems and services.

6. References

Alemayehu, N. (2003). Analysing performance variation using query expansion. *Journal of the American Society for Information Science and Technology* 54(5), 379-391.

Balzer, W., Moulines, C.-U. and Sneed, J.D. (1987). An Architectonic for Science, The Structuralist Program. Dordrecht: Reidel.

Barzilay, R. and Elhadad, M. (1997). Using Lexical Chains for Text Summarization. Proceedings of the Intelligent Scalable Text Summarization Workshop, ACL, Madrid.

Beaulieu, M., Robertson, S. and Rasmussen, E. (1996). Evaluating interactive systems in TREC. *Journal of the American Society for Information Science* 47(1), 85-94.

Belkin, N. J. (1993). Interaction with texts: information retrieval as information-seeking behavior. In *Information Retrieval ´93* (pp. 55-66). Konstanz: Universitetsverlag Konstanz.

Blair, D. (1990). Language and representation in information retrieval. Amsterdam: Elsevier.

Borgman, C. L., Hirsh, S. G., and Hiller, J. (1996). Rethinking online monitoring methods for information retrieval systems: from search product to search process. *Journal of the American Society for Information Science,* 47, 568-583.

Brajnik, G., Mizzarro, S. and Tasso, C. (1996). Evaluating user interfaces to information retrieval systems. A case study. In Proceedings of the SIGIR'96. ACM, New York, 128-136.

Bunge, M. (1967). *Scientific research.* Heidelberg: Springer-Verlag.

Burgin, R. (1992). Variations in relevance judgements and the evaluation of retrieval performance. *Information Processing & Management, 28*(5), 619–627.

Byström, K. (1999). *Task complexity, information types and information sources. Examination of relationships*. Doctoral dissertation. University of Tampere. Acta Universitatis Tamperensis 688.

Byström, K. and Järvelin, K. (1995). Task complexity affects information seeking and use. *Information Processing & Management,* 31, 191-213.

Carter, R. (2000). *Vocabulary. Applied linguistic perspectives*. 2nd ed. London: Routledge.

Chung, Y. and Lee, J. (2004). Optimization of some factors affecting the performance of query expansion. *Information Processing & Management* 40(6), 891-917.

Cooper, W.S. (1971). A definition of relevance for information retrieval. *Information Storage and Retrieval, 7*(1):,19-37.

Cosijn, E. and Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management, 36*, 533–550.

Dumais, S., Cuttrell, E., Chen, H. (2000). Bringing order to the Web: automatically categorizing search results. Proceedings of the SIGCHI conference on Human factors in computing systems 2000. The Hague, 145-152.

Ellis, D. (1989). A behavioural approach to information retrieval system design. *Journal of Documentation,* 45: 171-212.

Fowkes, H., and Beaulieu, M. (2000). Interactive searching behavior: Okapi experiment for TREC-8. In Proceedings of the BCS-IRSG: 22nd Annual Colloquium on Information Retrieval Research. Cambridge, 47-56.

Georgakopoulou, A. and Goutsos, D. (2001). *Discourse analysis*. Edinburgh: Edinburgh UP.

Hersh, W., Pentecost, J., and Hickam, D. (1996). A task-oriented approach to information retrieval evaluation. *Journal of the American Society for Information Science*, 47(1), 50-56.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation,* 52(1), 3-50.

Ingwersen, P. and Järvelin, K. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context.* Research monograph. Springer/Kluwer. 500 p.

Isenberg, D.J. (1986). Thinking and managing: A verbal protocol analysis of managerial problem solving. *Academy of Management Journal, 29*(4):,775-788.

Järvelin, K. and Ingwersen, P. (2004). Information Seeking Research Needs Extension toward Tasks and Technology *Information Research*, 10(1) chapter 212 [Available at http://InformationR.net/ir/10-1/chapter212.html]

Järvelin, K. and Kekäläinen, J. (2002). Cumulated Gain-based Evaluation of IR Techniques. *ACM Transactions on Information Systems (ACM TOIS), 20*(4), 422-446.

Joho, H., Sanderson, M. and Beaulieu, M. (2004). A study of user interaction with a concept-based interactive query expansion support tool. In: Proceedings of the ECIR 2004. Berlin and Heidelberg: Springer, 42-56.

Jones, S., Gatford, M., Robertson, S., Hancock-Beaulieu, M. and Secker, J. (1995). Interactive thesaurus navigation: Intelligence rules OK*? Journal of the American Society for Information Scienc*e, 46 (1), 52-59.

Kekäläinen, J. and Järvelin, K. (2002). Evaluating Information Retrieval Systems under the Challenges of Interaction and Multi-Dimensional Dynamic Relevance. In: Bruce, H., Fidel, R., Ingwersen, P. and Vakkari, P. (2002) (eds.), *Emerging Frameworks and Methods. Proceedings of the 4th International Conference on Conceptions of Library and Information Science (CoLIS 4),* July 21-25, 2002, Seattle, USA. Greenwood Village, CO: Libraries Unlimited, 253 – 270.

Koenemann, J. and Belkin, N. (1996). A case for interaction: A study of interactive information retrieval behavior and effectiveness. In*: Proceedings of the Human Factors in Computing Systems Conference (CHI'96)*. New York, NY: ACM Press, 205-212.

Kuhlthau, C. C. (1993). *Seeking Meaning*. Norwood, N.J.: Ablex.

Kuhlthau, C. C. (2004). *Seeking Meaning*. 2nd Ed. Westport: Libraries Unlimited

Kuokkanen, M. (1992), The Continuity Problem of Scientific Theories. An Example of Social Psychological Balance Theorizing. In H. Westmeyer (Ed.), *The Structuralist Program in Psychology: Foundations and Applications*. Seattle andToronto and Bern and Göttingen: Hogrete and Huber Publishers.

Laitinen, H. (1996). Reaalikoevastauksia rompulta: Tutkimus elektronisen oppimateriaalikokoelman käytöstä koevastauksien tuottamisessa [Answering high school exams with the help of CD-ROM: An investigation on using an electronic learning material collection when answering exam questions]. Tampere, Finland: University of Tampere, Dept. of Information Studies, M.Soc.Sc. Thesis, 78 pp. [In Finnish]

Morris, P. and Hirst, G. (1991). Lexical cohesion computated by thesaural relations as an indicator of the structure of the text. *Computational Linguistics, 17*(1), 21-48.

Over, P. (2001). The TREC interactive track: an annotated bibliography. *Information Processing & Management,* 37(3), 368-381.

Pennanen, M. and Vakkari, P. (2003). Students' conceptual structure, search process, and outcome while preparing a research proposal. *Journal of the American Society for Information Science and Technology* 54 (8), 759-770.

Pharo, N. (2004). A new model of information behaviour based on the Search Situation Transition schema. *Information Research*, 10(1) chapter 203 [Available at http://InformationR.net/ir/10-1/chapter203.html].

Qiu, Y. and Frei, H. (1993). Concept based query expansion. In: *Proceedings of the ACM SIGIR 16th International conference*, ACM: New York, 160-169.

Ruthven, I., Lalmas, M., and van Rijsbergen, K. (2003). Incorporating user search behavior into relevance feedback. *Journal of the American Society for Information Science and Technology* 54(6), 529-549.

Saracevic, T. (1976). Relevance in information science. In: M. Voigt and M. Harris (Eds.) *Advances in Librarianship*, vol. 6. Academic Press: New York, 79-138.

Saracevic, T. and Kantor, P. (1988). A study of information seeking and retrieving. Part II: users, questions, and effectiveness. *Journal of the American Society for Information Science,* 39, 177-176.

Schamber, L. (1994). Relevance and information behavior. In M. Williams (Ed.), *Annual review of information science and technology* 29 (pp. 3-48). Medford, NJ: Information Today.

Sihvonen, A. and Vakkari, P. (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. *Journal of Documentation* 60(6), 673-690.

Sormunen, E. (2002). Liberal relevance criteria of TREC—Counting on negligible documents? In: Beaulieu, M., Baeza-Yates, R. Myaeng, S. H. and Järvelin, K. (eds.) *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York, NY: ACM, 324–330.

Stairmand, M. (1997). Textual context analysis for information retrieval. In: *Proceedings of the ACM SIGIR 20th International conference*, ACM: New York, 140-147

Stokes, N. (2004). *Applications of lexical cohesion analysis in the topic detection and tracking domain*. Doctoral dissertation. National University of Ireland, Dublin.

Stubbs, M. (2001). *Words and phrases. Corpus studies in lexical semantics*. Blackwells: Oxford.

Suomela, S. and Kekäläinen, J. (2005). Ontology as Search Tool: A study of real users' query formulation with and without conceptual support. In: Losada, D.E. et al. (Eds.), Proceedings of the 27th European Conference on Information Retrieval, 21-23 March, 2005, Santiago de Compostella. Forthcoming

Vakkari, P. (1997). Information seeking in context; a challenging metatheory. In: Vakkari, P., Savolainen, S. and Dervin, B. (Eds) Information seeking in context. *Proceedings of an international conference on research on information needs, seeking and use in different contexts,* 14-16 August 1996, Tampere, Finland. London: Taylor Graham, 451-464.

Vakkari, P. (1998). Growth of theories on information seeking: An analysis of growth of a theoretical research program on the relation between task complexity and information seeking. *Information Processing & Management, 34*: 361-382.

Vakkari, P. (2001), A Theory of the task-based information retrieval. *Journal of Documentation, 57*(1), 44-60.

Vakkari, P. (2003). Task based information searching. In: Cronin, B. (ed.) *Annual Review of Information Science and Technology: Volume 37.* Medford, NJ: Information Today: pp. 413-464.

Vakkari, P. and Hakala, N. (2000), Changes in Relevance Criteria and Problem Stages in Task Performance. *Journal of Documentation*, 56(5): 540-562.

Vakkari, P. and Sormunen, E. (2004) The influence of relevance levels on the effectiveness of interactive information retrieval. *Journal of the American Society for Information Science* , 55: 963-969.

Vorhees, E. (2000), Variations in relevance judgements and the measurement of retrieval effectiveness. *Information Processing & Management* 36, 697-716.

van Rijsbergen, C. (1980). *Information retrieval*. 2nd ed. Butterworths: London.

Wagner, D. and Berger, J. (1985). Do sociological theories grow? *American Journal of Sociology, 90*: 697-728.

Wang, P. and Soergel, D. (1998). A cognitive model of document use during a research project. Study I. Document selection. *Journal of the American Society for Information Science, 49*, 115-133.

Wersig, G. (1973). *Informationssoziologie: Hinweise zu einem informationswissenschaftlichen Teilbereich.* Frankfurt, BRD: Athenäum Fischer.