Authors:	Järvelin Kalervo
Name of article:	An Analysis of Two Approaches in Information Retrieval: From Frameworks to Study Designs
Year of publication:	2007
Name of journal:	Journal of the American Society for Information Science and Technology
Volume:	58
Number of issue:	7
Pages:	971-986
ISSN:	1532-2882
Discipline:	Natural sciences / Computer and information sciences
Language:	en
School/Other Unit:	School of Information Sciences

URN: http://urn.fi/urn:nbn:uta-3-756
DOI: http://dx.doi.org/10.1002/asi.20589

[This is a preprint of an article accepted for publication in JASIST © 2006 Wiley Periodicals, Inc.]

An Analysis of Two Approaches in Information Retrieval: From Frameworks to Study Designs

Kalervo Järvelin

Department of Information Studies

University of Tampere – FINLAND

Email: kalervo.jarvelin@uta.fi

Tel: +358 3 3551 6953; Fax: +358 3 3551 6560

ABSTRACT

There is a well-known gap between *Systems-oriented IR* and *User-oriented IR,* which *Cognitive IR* seeks to bridge. It is therefore interesting to analyze approaches at the level of frameworks, models, and study designs. This article is an exercise in such an analysis, focusing on two significant approaches to IR, the Lab IR approach and Ingwersen's Cognitive IR approach. It focuses on their research frameworks, models, hypotheses, laws and theories, study designs and possible contributions. The two approaches are found quite different. This becomes apparent in the use of independent, controlled and dependent variables in the study designs of each approach. Thus each approach is capable of contributing very differently to understanding and developing information access. The article also discusses integrating the approaches at the study design level.

Keywords

Theoretical analysis; methodology; study designs; Laboratory IR; Cognitive IR

INTRODUCTION

In his acceptance address for the Salton Award for Excellence in Research at the ACM SIGIR 1997 Conference, Saracevic (1997) discussed the gap between the *systems-centered* (laboratory) and *user- centered* approaches to information retrieval (IR). The two communities do not recognize each other, do not interact, and as a rule, in systems-oriented projects, humans or users are absent. Ingwersen (1996) states that the communities are critical at each other, the latter suspecting the realism of the former, and the former suspecting the usefulness of the latter. Saracevic saw losing sight of human users as the greatest danger information science is facing. The two approaches would need to work cooperatively.

Ingwersen (1996) summarizes research related to the *Cognitive Approach to IR* and aims at developing cognitive IR theory. His article is also interesting in its discussion of the limitations of the Lab IR and User-oriented IR approaches. According to Ingwersen, the goals of the cognitive approach are to improve the intellectual access to information sources and, simultaneously, to provide IR systems with enriched contextual descriptions of the users and their situations that can better support users' information seeking. This goal is compatible with the goal of Lab IR. However, the cognitive research landscape is fundamentally broader, a sociological one (Ingwersen, 1996).

Since the year 2000, the Call for Papers for the ACM SIGIR Conference has not mentioned *information retrieval (IR) theory* as one of the key areas for which submissions are called. In 2000, papers were called for "IR Theory, including logical, statistical and interactive IR models, and data fusion", among others. Since then, the corresponding item in the Call for Papers has been, more often than not, "Formal Models, Language Models, Fusion/Combination", which also suggests how *theory* is to be interpreted in the SIGIR context. Studies on interaction or users belong under another heading, typically consisting of the subheadings interactive IR, user interfaces, visualization, user studies, and user models.

This article is about IR theory in a different sense. It looks at two significant IR approaches as examples and purports to analyze, what one may learn about information retrieval or say about information retrieval by employing such approaches. We shall analyze the mainstream Lab IR and Ingwersen's Cognitive IR approaches. The Lab IR approach was chosen as one example because it to a large degree defines IR – at least in the ACM SIGIR context. Ingwersen's Cognitive IR approach (1996) was chosen as the other example since it provides a fairly recent comprehensive framework for IR, and has generated empirical research explicitly based on it. It explicitly seeks to bridge the gap between systems and user orientation.

Aims and Focus. In this article we want to analyze the differences between the approaches down to the level of study designs where variables are defined and hypotheses stated. Any concrete integration must exemplify itself in the use of variables and hypotheses. The approaches to research in IR have a major role in guiding the study designs. Therefore this article addresses the following questions:

- What is the representative conceptual *framework* of the approach under scrutiny? Which phenomena are suggested as important to study? Which are not?
- What is the *model* of the approach? How does it represent and relate the phenomena to be studied? Where is the focus, what are fringe areas, what is excluded?
- What kind of *hypotheses, laws* and *theories* may one test within each approach?
- What kind of *research designs* are proposed and *contributions* offered?

The concepts in italics above are clarified in the third section on conceptual tools. A thorough analysis of this kind cannot be covered in much detail in one article. Therefore the paper is an *exercise* in the analysis of frameworks, models, theories and study designs in IR – or in the aboutness of IR. This means that examples within the two approaches are drawn to analysis and discussion. The goals are also more modest than in Philosophy of Science. We do not aim at analyzing the evolution of 'theories' of IR from the Philosophy of Science viewpoint, just at an analytical exercise focusing on two relevant cases. Similar analysis is possible regarding other IR approaches, be they systems or user-centered or cognitive.

Approach and Methodology. There is no distinguished name for the method employed in this article. As for the frameworks and models, this article analyses the two examples in IR for their content (objects and relationships of interest) and their ramifications. Regarding hypotheses and theories, the article is reconstructive. It tries to explicate (reconstruct), which kinds of hypotheses and theories are possible in each approach. As for research designs, it seeks to outline the types of designs the approaches have led to, or support.

Contributions and Limitations. This article seeks to contribute, technically, (a) an approach and concepts for analyzing theoretical and study design aspects of IR, (b) analytical statements about the scope and features of two notable IR approaches, (c) analyses of possible research questions in the two approaches, and (d) recommendations for extension of IR approaches if further integration is desired. This article does not propose nor evaluate any IR techniques. It neither assesses IR research directly nor proposes any given line of research to be followed. We try to establish facts about the approaches under examination, not present opinions. To state that an IR approach contains or lacks a given feature, is not an opinion, but a verifiable statement. To say that an approach should cover a given feature is an opinion. The latter remain in the reader's domain. Yet a qualified statement, saying 'in order to achieve X, this approach should contain the feature Y' is, again, a verifiable statement (being instrumental knowledge, if true). While avoiding opinions, this paper seeks to facilitate discussion on what IR is about. We may approach IR in one way, or another; however, whatever we choose has consequences on what IR research is and may contribute to information access.

Terminological Note. We shall use "Information Retrieval" (IR) for the discipline and "information retrieval" or "information access" for real-life acts aimed at information access through retrieval systems.

The next section discusses approaches to IR research and prior analyses of this research. The following section discusses concepts used to analyze the two approaches. We shall then analyze the Lab IR approach to IR in the 4th section and the Cognitive IR approach in the 5th section. Sections for discussion and conclusions end the article.

PRIOR ANALYSES OF IR APPROACHES

Approaches in IR

There are several introductions to approaches in IR research. Ingwersen (1992) reviews the three major approaches: Lab IR, User-oriented IR and Cognitive IR approaches and Ingwersen and Willett (1995) provide an introduction to systems-oriented and cognitive approaches to IR. Belkin (1990) and Allen (1991) review the Cognitive IR approach. Ingwersen and Järvelin (2005) discuss all three approaches. We shall focus on the Lab IR and Cognitive IR approaches here; the user-oriented one is subsumed by the cognitive one in the 1990's.

The Laboratory (systems-oriented) IR approach. Systems-oriented IR research may be divided into studies in operational IR environments and studies in laboratory IR environments (e.g., Ingwersen & Järvelin 2005, Section 4). We focus on the latter which represents a great majority of systems-oriented IR research. While Lab IR encompasses many special domains – e.g., ad hoc retrieval, text summarization, question answering, filtering, clustering and classification, cross-language retrieval, topic detection and tracking, text mining, speech, music, image and video retrieval, hypermedia retrieval, structured document retrieval – there is high consensus regarding the approach to study. The specific study designs and effectiveness measures of course vary by sub-domain. Nevertheless, the test-collection based approach to ad hoc text retrieval (or Cranfield approach) directly represents much of IR research and serves as a paradigm for many of the remaining areas. We call this approach the Lab IR approach and analyze it in a dedicated section below.

Interactive IR involves, at least in a limited way, interactive users in study designs (e.g., the TREC interactive track, Hersh & Over, 2000). Also modern Web IR research, based on web query logs and click-through data, at least indirectly involves some user aspects. We shall return to these in the discussion section.

The Cognitive IR approaches. Cognitive and user-oriented IR research developed from 1970s onwards – see Ingwersen and Järvelin (2005, 195-203) for a recent discussion. When Lab IR research could be seen to neglect information seekers in its modeling and experimentation, cognitive IR research focused precisely on them. Among early approaches we have Fidel and Soergel (1983) presenting factors affecting online searching, and cognitive models of interaction by Ingwersen (1982) and Belkin (1984; Belkin *et al.,* 1982). More recently, Kuhlthau (1993) developed a six-stage model of information seeking, which has inspired other researchers concerned with interactive IR in the 1990s. For example, Vakkari (2001) extended Kuhlthau's approach to task-based IR, founded on longitudinal empirical studies. The model related stages in work task and search task performance, the actor's knowledge, and relevance assessments with each other. Belkin and colleagues (1993; 1995) categorized information seeking strategies (ISS) into a comprehensive multi-dimensional model of ISS behaviors or 'episodes'. The fundamental idea is that people engage in multiple searching behaviors during IR sessions as well as across sessions. Ingwersen (1992; 1996) further elaborated his early work on the cognitive approach in the 1990's toward a comprehensive research program and a theory of Cognitive IR.

Saracevic's stratified model of interaction levels (1996) - surface processing level, cognitive communication level, situational level, and affective level – also placed users in a situation in a context. He also contributed several types of relevance – algorithmic, topical, pertinent, situational, and emotional/intentional – involved in information access. Wang and Soergel (1998) presented a document selection framework as a stage model. It focuses in great detail on the decision stages on document selection/rejection/use by applying several relevance criteria and document value dimensions. Pejtersen and Fidel (1999) proposed a multi-dimensional design and evaluation framework for cognitive systems engineering and work analysis, with applications to IR systems design and evaluation. Wilson (1999) presented a nested model of information behavior covering information retrieval as an aspect of information seeking.

As the discussion above suggests, there are multiple approaches to Cognitive IR and no consensus comparable to Lab IR regarding the approach – one has to choose one or more for analysis. The approaches exhibit different levels of detail, different focus and coverage, different conceptual breakdowns and relationships.

We have selected Ingwersen's approach (1996) as the representative of, and target of analysis among, cognitive IR approaches. His approach aims to broadly cover information access from tasks to retrieval and information use, it is explicit and analytical about the stakeholders and the cognitive structures associated with them, and aims at generality as opposed to special cases (cf. Järvelin & Wilson, 2003). We call this approach (Ingwersen's) Cognitive approach and analyze it in a dedicated section below.

As the result of our selections we have one popular framework, covering much or Lab IR research, and another much less popular regarding the number of studies or researchers, as an example of cognitive IR approaches. This is unavoidable because in the latter area there is no single dominant approach. However, Ingwersen's Cognitive approach a good representative of the cognitive approaches for analysis.

Some Prior Analyses of IR Approaches and Methodologies

Tefko Saracevic has critically discussed evaluation in IR research (1995), presented a stratified model of IR interaction with an associated system of relevancies (1996), and called for the integration of user-oriented and system-oriented IR research (1997). He argues that much of IR evaluation focuses on the processing level (evaluation of algorithms) while providing minor streams of evaluation at the levels of users and uses, markets and products, and social impacts. A major problem is the isolation of evaluations at different levels, even lack of interest at the processing level toward other levels of evaluation. Saracevic also criticizes the sole use of relevance, recall and precision as measures in evaluation and calls for proper measures at the levels of users and uses, markets and products, and social impacts. However, he does not make a specific study design proposal for the incorporation of users and context in IR. Part of such work may still be read in the large empirical study (Saracevic & Kantor, 1988) which incorporated several user-related variables.

David Ellis (1996) discusses the Lab IR approach and the Cognitive IR approach as paradigms and finds both possessing only a fragmentary or ambiguous paradigmatic identity. The former is stronger in its symbolic generalization (e.g. standard measures), metaphysical assumptions about the constitution and purposes of IR systems, and sociological commitment to quantification hopefully leading to exact science. What is lacking is an exemplar, a model problem solution, which could be transferred to other IR problems. The latter, Cognitive IR, however, is weaker in many paradigmatic features while the distributed expert IR systems approach might serve as an exemplar. This paradigmatic weakness motivates our view on choosing one example of cognitive approaches (Ingwersen's) under study.

Ellis (1996) also discusses the Lab IR approach and the Cognitive IR approach as *research traditions*. A research tradition has a metaphysical and methodological character which differentiates it from other traditions; it has a problem-determining role, a heuristic role and a justificatory role. The Lab IR approach clearly has a strong research tradition. Nevertheless, Ellis questions the empirical problem-solving power of the lab approach due to validity problems of the performance evaluation (relevance, recall, precision) and the applicability of the results to operational systems. The same conceptual problems have plagued the approach right from the beginning. One may say that Lab IR abstracts a mechanical component out of human–literature communication and making it the (isolated) object of inquiry. Removing individual subjectivity fosters homogeneous quantitative treatment of IR systems – but at the cost of not being able to handle problems raised by human behavior and capabilities.

According to Ellis, Cognitive IR also has many characteristics of a research tradition. However, the approach suffers from measurement problems at the linguistic (relevance) and cognitive levels (changes in knowledge states/structures) of interaction, not to mention wider philosophical issues related to knowledge structures and their change, subjective and objective knowledge, etc. Cognitive IR explicitly seeks to model cognitive states as the basis of retrieval but at the cost of avoiding complex relationships between knowledge structures and their textual representations. We think that this analysis is pertinent, but seek to analyze in this paper, what kinds of research designs the two approaches foster, allow or inhibit based on the kinds of variables they propose for the designs.

Peter Ingwersen (1992; 1996) discusses the user and lab-oriented approaches to IR and points out their limitations and the communication gaps between them. He proposes his Cognitive IR approach, discussed in detail below, as a general one, subsuming both the user-oriented and the lab-oriented one. Borlund (2000) follows this approach and analyzes the tradition of IR evaluation. Lab evaluations are found wanting in realism and a new evaluation package for interactive IR is proposed. Genuine user-based relevance assessments are found necessary. However, truly user-based evaluation is costly, suffers from problems in control (repeatability) and result comparison. Kekäläinen and Järvelin (2002) examine the rationale of evaluating the IR algorithms, the status of lab evaluation, and the applicability of non-traditional evaluation methods and concepts, such as contextual relevance concepts (e.g., Saracevic, 1996; Cosijn & Ingwersen, 2000). They find the Lab approach limited but viable for the specific task of evaluating IR algorithms. The critics of the model are right in their claims but this rather suggests additional, broader evaluation scenarios than discarding the old one. Hjörland (2005) analyzes empiricism, rationalism and positivism in Information Science. This work, however, remains at a philosophical level and is not specific to IR study design.

Tague (1992) discusses the pragmatics of IR experimentation in terms of 10 decisions that the investigator has to consider, including the type of test (lab, operational), the definition and operationalization of variables (e.g., dependent, independent), experimental design, and data collection and analysis. The sample variables discussed in the paper are quite comprehensive regarding the information retrieval process, some covering systems aspects and others user / searcher aspects. Therefore Tague's work may serve as a methodological guide for both Lab IR and Cognitive IR approaches. However, her work does not aim at proposing an approach or particular study designs for IR.

In this paper, we shall remain at a general level as well but in a different way than in the research reviewed above. We want to analyze the path from the approaches to research designs and find out what kinds of designs the approaches suggest. How well defined are they, how may they be elaborated? What sound possibilities are there?

CONCEPTUAL TOOLS

In this section we clarify the concepts of 'framework', 'model', 'hypothesis', 'law', 'theory', 'study types' and 'contribution'. They are used in the two ensuing sections to analyze the approaches to IR research. An approach is seen to contain a framework and possibly some of the other components, perhaps in multiple versions.

Frameworks in Research. All research has an underlying model of the phenomena it investigates, tacitly assumed or explicit. Such models, called conceptual frameworks (Engelbart, 1962), paradigms, conceptual models or just models (Wilson, 1999), often become topics of discussion when the orientation of a research area is debated. Defining conceptual frameworks means, according to D.C. Engelbart (1962), specifying:

- Essential objects or components of the system to be studied.
- The relationships of the objects that are recognized.
- The changes in the objects or their relationships that affect the functioning of the system.

- Promising or fruitful goals and methods of research.

The terminologies may vary but still there are shared *ontological* (what is out there to investigate?), *conceptual* (how to name that?), *factual* (what to take as givens?), *epistemological* (what can we learn about?), and *methodological* (how can we learn about it?) assumptions in research areas.

Often frameworks are implicit, unarticulated, yet socially shared in a research community. Dervin (1999) notes that frameworks can be used to release research from implicit assumptions and draw them to daylight for examination. This is important for the present discussion.

Frameworks are broader and more fundamental than scientific theories in that they set the preconditions of theory formulation. In fact, they provide the conceptual and methodological tools for formulating research questions, hypotheses and theories. Therefore they cannot be assessed directly empirically but only in terms of their problem-determining, heuristic role and justificatory roles – the quality of research they create. They are necessary for scientific growth – not much knowledge cumulates through totally individual approaches.

Models. The word ‘model’ has many meanings even in academic disciplines. It often appears in texts of IR in the form of ‘formal model’. Formal models in Logic, Mathematics or IR do not state anything about the real world but rather consist of analytical statements relating mathematical constructs to each other. Still one may think that a formal model in IR is an abstract representation of a real-world IR system and thus also a theory on how it functions. Empirical sciences contribute theories (see below) that have real testable (refutable) content, as does IR evaluation.

In this paper we define the concept *model* to refer to a precise (often formal) representation of objects and relationships (or processes) within a framework. In principle, modeling may also involve modeling human actors and organizations as well.

Hypotheses and Laws. A hypothesis states a verifiable fact whose truth is unknown. A hypothesis needs to explain the facts or regularities that gave rise to the study; it must be logically non-contradictory and precise, testable in principle, informative and simple. Further, a hypothesis is only accepted provided that it passes rigorous tests. Some hypotheses state an expected relationship through abstract concepts (theoretical hypotheses). Other hypotheses are grounded on specific datasets, measurements and procedures (operationalized hypotheses). It remains the skill of the researcher to operationalize theoretical hypotheses into operational ones for testing with available data.

The concept of scientific law is difficult – telling the difference between law-like and random generalizations. Empirical scientific laws express verified relationships between observable objects, properties or events. Theoretical laws refer to non-observable objects or properties. Laws may express deterministic or probabilistic regularities (e.g. Zipf's Law).

Theories. Laws are the building blocs of theories. The difference between laws and theories is that theories typically consist of systematic collections of theoretical and empirical laws and associated existence assumptions. A theory explains observed regularities and hypothesizes novel ones. Further, a theory provides deeper understanding of phenomena by using theoretical concepts that go beyond immediate observations. Therefore scientific theories represent reality, systematize knowledge concerning it, and guide research, e.g., by suggesting novel hypotheses (Bunge, 1967).

Theoretical growth in a research area may incur from theory expansion (e.g., due to enrichment through new concepts), greater analytical power (through formalization or model building), improved empirical support (due to confirmed hypotheses), and proliferation of hypotheses within the theory (Wagner *et al.,* 1992). Conceptual development may be due to increased preciseness, accuracy, simplicity, generality, and ability to suggest hypotheses of concepts (e.g., Cohen & Nagel, 1936). Theoretically good concepts relate to each other in systematic and fruitful ways. What kind of theoretical growth is there in the two IR approaches?

Study types, designs and contributions. The discussion above mainly relates to empirical studies, but there are other types as well. Studies may be classified into (a) *empirical* (descriptive, evaluative, explanatory) studies, (b) *theoretical* or conceptual (theoretical, metatheoretical), (c) *methodological* (creating and analyzing methods), and (d) *constructive* (designing and testing systems). Likewise, the contribution types of each study type are:

(a) Empirical: novel observed laws, confirmations or refutations of hypotheses.

(b) Theoretical: novel or elaborated theories, frameworks, concepts, or models.

(c) Methodological: novel or better methods, application guidelines, or analyses of pitfalls.

(d) Constructive: novel, better systems or practices.

In empirical studies the *study design* specifies a set of research questions, some methods of analysis and a data set. Research questions are expressed within a given framework and may apply some model or theory. In a study design, one may specify the interaction of several types of variables to be examined (e.g., Tague, 1992):

- *Dependent* variables – the variation of which is explained.
- *Independent* variables – the ones systematically varied in order to see the responses in the dependent ones.
- *Concomitant* variables – the ones fixed to prevent uncontrolled variation in the results.

The scales of measurement for these variable types are the *nominal, ordinal, interval,* and *ratio* scales. Other variables remain as *hidden* variables.

THE LABORATORY APPROACH TO IR

The framework presented below is the general laboratory IR framework shared in many studies. This does not mean that there would not be exceptions not fitting into it. While there are subareas of systems-oriented IR that do not follow this framework in detail, they are equally systems-oriented and non-user-oriented so that the present discussion is relevant regarding them as well.

The Basic Framework

The basic laboratory IR framework is depicted in Figure 1. It has no human searcher involvement but a 'user' could be fitted into the right-hand feedback arrow (interactive lab experiments). However, this framework does not allow a 'user' as the initiator of the 'topic' since all topics are predefined. It suggests documents, search requests, their representation, the database, queries, and the matching of the latter two as foci of research. Methodologically, it also suggests recall base construction (relevance assessments) and evaluation as foci.

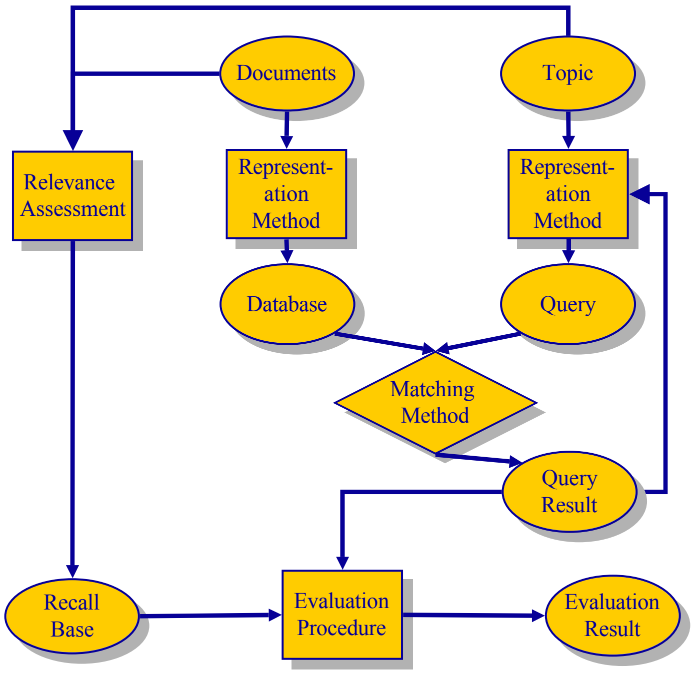

Figure 1. A portrait of the laboratory framework to IR

We shall analyze below how the objects and processes of Figure 1 are typically understood in the framework and thereafter which possible objects or processes are excluded.

Documents, Collections and Databases. Documents are natural language texts without structural mark-up other than phrasal, sentential and paragraph structure.[1] The texts consist of words belonging to some natural language, may have a part-of-speech (POS) and position in the text. Documents have a known length. The words serve as the source of indexing features. Document collections are sets of independent documents with collection size as an important attribute. Collections have domains sometimes used to classify them but hardly ever used as an explanatory feature.

Documents are represented through independent indexing features derived from their words. The features are derived from word tokens by a representation process, which may involve

[1] Text is used as an example here, representing all kinds of information bearing units.

morphological normalization (e.g., stemming) and other NLP. The features have a within-document frequency and a collection frequency and, through their positions, co-occurrence. A database is a set of documents with feature-based representations somehow organized for efficient retrieval (e.g., through an index structure).

There are a number of document features that are seldom used in the framework. These include document genre, disciplinary or organizational (working life / leisure) domain, inter-document dependencies, or discourse. Likewise, document representation typically means full content representation, not structure or metadata representation although there is recent work in structure-based representation (e.g., the INEX campaign, http://inex.is.informatik.uni-duisburg.de).

Requests and Queries. Requests reflecting information needs are natural language texts without structural mark-up other than phrasal, sentential and paragraph structure. Again, these texts consist of words of some natural language, and may have a part-of-speech (POS) and position in text. The words serve as the source of indexing features. Sometimes, e.g., in TREC, requests have three components: title, description and narrative. This supports the study on the effects of request length. In Lab IR experiments, the requests are called topics. The topics are purely topical (vs. factual), content-only (vs. metadata based), well-defined (vs. vague), static (vs. dynamic) and exhaustive (vs. high-precision oriented).

Like documents, requests are represented through independent indexing features derived from their words perhaps using NLP. The features have a within-request frequency or weight. While unstructured (bag-of-words) queries are most typical, queries may be structured as well. Queries may be expanded through some automatic or interactive expansion method.

While query length has been used as an explanatory feature, there are other request/query features that have seldom been.[2] These include request complexity (e.g. the number of logical facets) and specificity (specific search keys), type of content search keys (e.g. proper names), and query goals (exhaustive vs. the single best document[3]; topical vs. verificative goal). Searcher's (un)certainty about the request/query has also been neglected. Real-life requests may also contain non-content access handles, like metadata, which mostly lack from the studies within the framework.

Matching and Query Results. Matching is based on document and requests representations as guided by a retrieval model. There is a lot of variety among retrieval models (see the subsection on models). This is the real focus of IR systems development and provides a nominal scale independent variable, i.e., the type of retrieval model.

Query results are typically ranked lists of document representations. The list items have as their properties a rank, score and (binary) relevance. The ranked lists have aggregate properties like length as well as recall and precision calculated as, e.g., uninterpolated MAP (mean average precision) or otherwise. These calculations depend on the recall base features. Buckley and Voorhees (2004) addressed the issue of incomplete relevance assessments.

Recall Base and Evaluation. The recall base is derived by extensive pooling of possibly relevant documents for each topic. The participants of test collection construction retrieve, for each topic, documents and the *Top-n* (*n* being, say 100) results are merged and then assessed by independent assessors typically using binary relevance and very liberal relevance criteria (Sormunen, 2002). Each pool is assumed to contain all relevant documents for a topic.

---

[2] Most recent work at TREC (http://trec.nist.gov/), INEX (http://inex.is.informatik.uni-duisburg.de/), CLEF (http://www.clef-campaign.org/) and NTCIR (http://research.nii.ac.jp/ntcir/) however cover some of these features.

[3] Voorhees (2001) analyzed retrieval effectiveness based on the best document.

Evaluation may use various metrics, typically based on recall and precision, e.g., MAP across a topic set.

Excluded Objects. Figure 1 has no explicit IR system interface although it can be seen included in request representation as a feedback feature. Interface functionality was vividly analyzed in the 1980's (e.g. Belkin *et al.,* 1987) but this line of research has died out.[4] The framework does not suggest this as a focus.

Much of Lab IR research disregards searchers. However, "users", while having no interesting explicable attributes, are nevertheless hiding in the relevance assessments as hidden variables. Relevance assessments are rarely seen as problematic nor essentially related to anything else than the requests and documents. Interactive Lab IR research takes searchers into account (e.g. the TREC Interactive Track; Hersh & Over, 2000). However, in these efforts, searchers are made to find (through a given a system) documents for given static topics that someone else assessed as topically relevant.

Excluded Processes. Figure 1 has just two processes of representation, one of matching and an external process of relevance assessment / evaluation. Regarding information retrieval, the search process is not represented nor the other information access processes of which retrieval forms a part and contributes to in practice. This may be significant since no subsystem of information access (say, information retrieval) is independent of the information access environment where systems and practices are used in a concerted way.

Motivations of the Framework. The Lab IR framework is a framework of the IR phenomenon *and* system evaluation. It supports experimental control of variables, sharing of research efforts, and comparability of results. Other objects or processes than those discussed above play negligible roles in research within the framework. However, it has not been shown that the excluded aspects would have a significant effect on the processes in focus, in particular on explaining the variation of recall and precision. Therefore the framework does not suggest how to look at them nor encourages this. They are not seen relevant.

The Model(s)

Lab IR has developed many retrieval models, e.g., the Boolean Model, the Vector Space Model, etc. A *retrieval model* consists of the specification of document and request representation, and of the definition of the matching algorithm for comparing these representations. In principle, *document representation* may concern the representation of the document content, layout, structure, and metadata. Most IR models focus on content representation. *Request representation* may concern the representation of request content and its target structure as well as request structure through query language operators.

Matching algorithms are formal methods for computing the similarity of query and document representations. Belkin and Croft (1987) classified them as *exact match methods* (e.g. Boolean logic) and *best match* methods, see Figure 2. The framework together with the models have been strong tools in facilitating a debate on the principles of modeling. Lab IR models only active *algorithmic* components in information retrieval.

---

[4] Also noted by Amit Singhal in his keynote address at ACM SIGIR 2005, Salvador, Brazil.

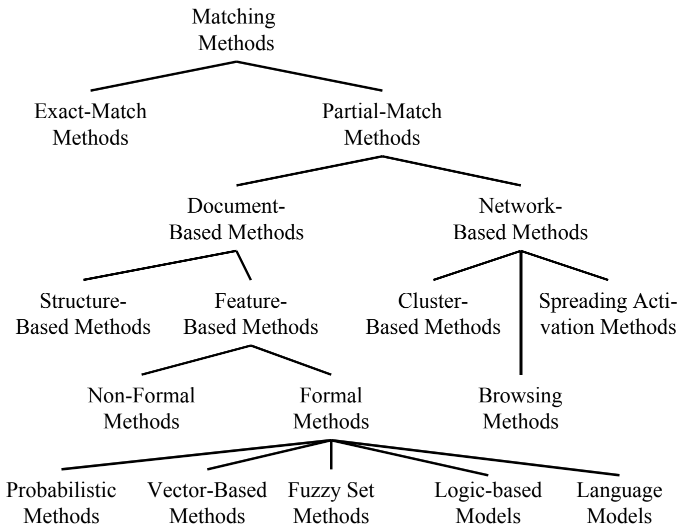

Figure 2. Matching algorithms (based on Belkin & Croft, 1987)

Hypotheses, Laws and Theories

Lab IR develops techniques for finding relevant documents. It is thus a branch of technology, not a branch of science. The capability of the techniques in finding relevant documents is usually measured in terms of recall and precision. Therefore Lab IR hypotheses, laws, and theories are about the explanation of the variation of recall and precision through various metrics. The independent variables typically are the use or non-use of various techniques developed within one or more retrieval models.

For example, in automatic indexing we may observe the Zipf's Law of word frequency distribution in corpora and apply Information Theory to analyze the signal value of words. These are utilized in automatic indexing theory, which is about document representation through its content words (perhaps after some NLP) and relates the within-document frequency (*tf*) and collection frequency (*df*) with recall and precision (see, e.g., Salton & McGill, 1983).

As an example of hypothesis-driven study, Buckley and Salton (1995) studied the optimization of relevance feedback weights of query keywords. Optimization starts by a query key weighting scheme based upon Rocchio feedback and then improves query key weights dynamically by testing possible changes of weights on a learning set of documents. Based on prior work, they formulated four conjectures[5] stating the relationships, e.g., between (1) retrieval performance and the magnitude of weight revisions, and (2) the contribution of the original request and the quality of the weights of the added keys. These two were confirmed, the two others not. Both conjectures relate request representation to retrieval effectiveness.

These examples fit into the foci of the Lab IR framework and the retrieval models in their selection of independent and dependent variables. Explanation through, e.g., a *different type* of document collection or request set is rare – if not coupled with a novel technique, such an explanation may not be considered a contribution.[6]

Robertson (2000) argues that IR is not a very theoretical field but rather pragmatic, driven by pragmatic problems and evaluated by practical criteria. There are few strong theories in IR and certainly no overall theory *of* IR. His discussion on the effects of 'precision devices' in set-based vs. ranked output IR systems is illuminating. While we may gain much insight through theoretical argumentation on whether recall and precision may be expected to change in a given direction in a ranked output system, we cannot make a well-founded hypothesis, but need to test the effect. This is due to hidden threshold changes in the systems.

---

[5] Conjectures may seem more like guesswork than hypotheses but are of the same type nevertheless (Websters 3rd New International Dictionary) and we do not know whether the authors intend there to be a difference.

[6] The central position of novel algorithms (techniques) stems from the fundamental underlying question of Computer Science – *What can be (efficiently) automated?*(Denning, 1997).

The models and the framework invite one to formulate hypotheses, establish laws and aim at theories that relate the focal components of retrieval models to retrieval effectiveness, that is, the use/non-use of techniques[7] for document and request representation and matching to recall/precision. They do not encourage, but neither deny, the exploration with various types of document collections, request types, or kinds of relevance assessment (e.g. Cosijn & Ingwersen, 2000). They do not help in this, either, since no analytical variables are proposed and thus nothing is seen out there.

Study Types, Designs and Contributions

Study Types and Contributions. The study types within the Lab IR framework are of all major types. The archetype of a Lab IR study is an *empirical* IR experiment where two or more IR techniques are described and tested for their recall–precision effects using some test collection and the well established IR test methodology. The contribution is a proposed novel technique and the finding whether it is effective. If one stresses the technique, these studies may also be called *constructive.* Another typical set of studies within the approach is *theoretical* studies, which tend to formally analyze and develop IR models. They contribute new ways of document / request representation and matching. Figure 2 shows many contributions over the years. Finally, there are *methodological* studies mainly focusing on experimental methodology (e.g. Hull, 1993). They have contributed novel effectiveness measures, guidelines for setting up test collections and experiments and assessing reliability and significance of findings. Below we shall focus on IR experiments.

[7] The expression "the use/non-use of techniques" is deliberate since a technique is not a variable, while its use/nonuse is, in a study design.

Experimental Study Designs. Lab IR study design is schematized in Figure 3 as a modification of Figure 1. The nodes with darker fill, *focus areas,* are given with variables, e.g., M_k for the matching method node. Nodes with lighter fill, *fringe areas,* do not have associated variables. This is used to suggest that the Lab IR framework invites one to design studies on the interaction of any combinations of given variables and disregard anything else. Therefore, in such study designs, the foci of elaboration typically are the retrieval models and their refinements – the two representations with variables R_i and T_j and their matching M_k, with their outcomes D_i, Q_j and the query result A_{ijk}. Another focus area is evaluation, where the evaluation procedure P_e covers alternative methods (such as recall based vs. DCV-based[8] measurement), metrics (such as uninterpolated MAP vs. *R*-precision) and tests (e.g. t-test or Friedman's test) that produce the evaluation results E_R.

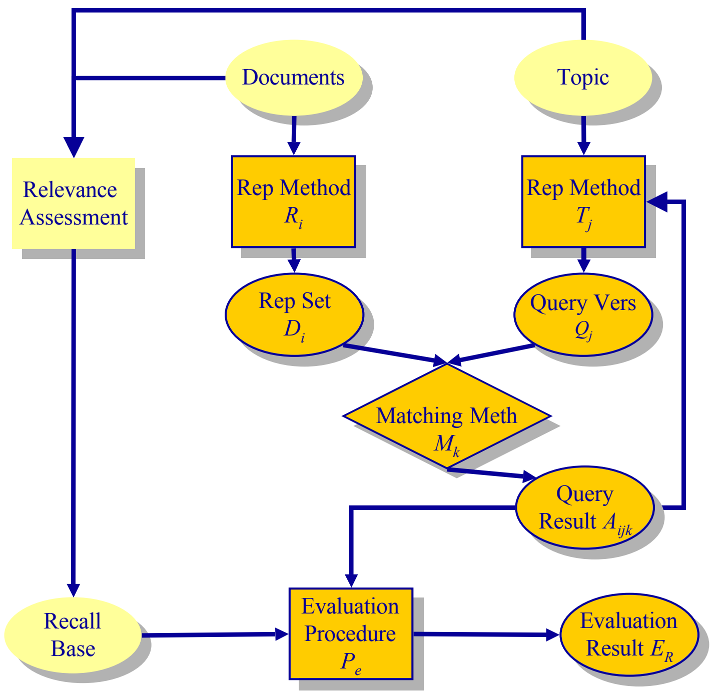

Figure 3. Laboratory IR study design schematized

[8] Document Cut-off Value based, i.e., based on predefined inspection points along the ranked result list.

In a study design, one may investigate the interaction of all combinations of the variables offered. Typically, some variables R_i, T_j and M_k are chosen as independent variables and the query result A_{ijk} as an intermediate dependent variable. For instance, language normalization N with values {'nothing', 'Porter stemmer', 'lemmatizer'} may be an independent (nominal scale) variable, each value representing the use of a particular technique. By applying some procedure of evaluation, the evaluation result E_R becomes the final dependent variable. The test collection (topics, documents, and recall base) is controlled by choosing a standard collection for the experiment. Other possible variables are hidden ones.

The *fringe areas,* topics, documents, relevance assessments, and recall bases, are seen as less problematic. They *do not contain interesting variables*. Indeed, topics and documents are seen as sources of indexing features and the focus is on their representation, not on their *kind.* Relevance assessments (binary, topical, static) and recall bases are fixed for each set of documents and topics and not based on anything else than topicality. Recall base and the topic set *sizes* are methodological concerns (e.g. Voorhees, 2001; Sanderson & Zobel, 2005).

Contribution Revisited. Theoretical growth in Lab IR research provides the four types introduced in the section on conceptual tools, but in a particular way, over the years. There is *theory expansion,* e.g., due to enrichment of the automatic indexing theory through document length normalization. There is *greater analytical power* through formal model building and more refined empirical methodology. Further, there is *improved empirical support* for several IR models through testing over several test collections (albeit of the same type). While there are reservations regarding the formulation of hypotheses (see above), Lab IR has constantly produced *new research questions* – even if only within the types of variables of Figure 3.

Theoretical growth in Lab IR has been constrained by not bringing radically new concepts or relations (and hence, associated variables) into the framework or the models. This constrains theory expansion to the domain of Figure 3. It also constrains the analytical power: if there were, e.g., established variables on types of requests, that would suggest new analytical relationships (and thus research questions) between request types and the dependent variables, mediated by the representation and matching variables. Moreover, even empirical support might be improved in a different way, if there were variables associated with test collections, leading to novel types of collections.

Whether some IR technique behaves differently, e.g., with radically different kinds document collections or request types, or searcher types, cannot be analyzed nor explained unless the framework is expanded by suitable variables. This remains a matter of belief until tested. If one is convinced of the generalizability of the current findings (on best techniques) over all reasonable large contexts of IR systems application, the issue does not matter and testing remains unlikely. If one is not, it does require testing. Any hidden variables however become operational when one transfers the results into the real world.

Finally, one may ask what is, or could be, explained though the experiments. Current Lab IR explains the variation of recall and precision. If one wants to understand also something else in information access, new concepts may be needed in the framework.[9]

INGWERSEN'S COGNITIVE APPROACH TO IR

The cognitive framework discussed here is the one proposed by Ingwersen (1996). His paper summarizes research related to the 'cognitive approach' and aims at developing cognitive IR theory. Further, the paper discusses research frameworks and methods, actual information access phenomena, and system design. However, we focus only on the framework/theory aspects.

The Framework

Figure 4 presents the Cognitive IR framework. The framework points to a number of objects: social/organizational environments, individual users, information objects, interfaces, requests and queries, and IR system settings as foci of analysis. These foci consist of *cognitive structures* of varying origins (of authors, indexers, system designers, etc.). The concept of cognitive structure remains vaguely defined but may be seen as consisting of concepts and their relationships related to, possessed by, or extracted from the objects.[10] Each object may have more or less detailed models of the other objects as pointed by the arrow-labeled 'Models' in the figure[11]. The unidirectional arrows represent cognitive transformations and influence, e.g., an author transforms his/her cognitive structures into an information object. The bidirectional arrows represent interactive communication of cognitive structures, e.g., an individual user transforms her deficient cognitive structures into a request for information. (Ingwersen, 1996)

---

[9] We are aware of the more constructive research in IR, seeking to explain efficiency of IR techniques, e.g., the MG system studies (Witten, Moffat & Bell, 1999).

[10] The cognitive structures (or knowledge structures) are about task domains, retrieval processes, etc. – i.e. knowledge about these, described at length by Ingwersen (1992). What the addition of the word 'structures' brings to this aboutness, is not rigorously defined (see also next subsection).

[11] Belkin (1984) discussed these mutual models as well.

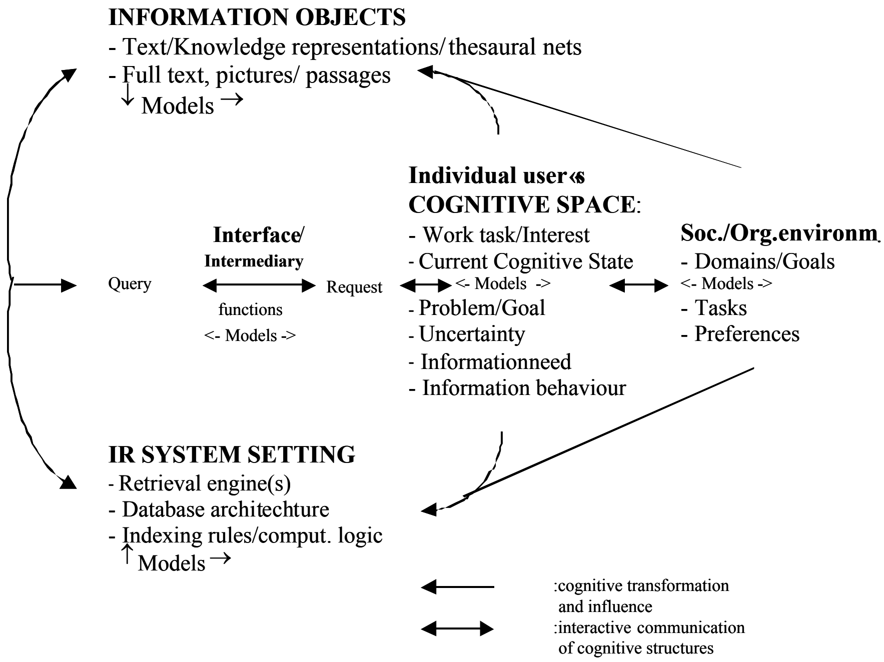

Figure 4. The cognitive IR framework (Ingwersen, 1996)

Comparing to the Lab IR framework, one notices new objects, properties and relationships of interest. Also the interface is made explicit. The Cognitive IR framework has much less research than the Lab IR framework done within it. Therefore it has less chance of being well defined and with an established methodology. It is nevertheless important to look how the objects and relationships are understood and represented (the subsection on models below) within it.

Cognitive Structures. The framework in Figure 4 suggests some cognitive structures to be observed. Some of them are computer software / data structures, some obviously text, and some mental. Their specific representations as cognitive structures are not defined, e.g., whether information object representations are any different from their content as text or whether or not goals or tasks are represented as plain text. Nevertheless, the framework suggests a number of objects and cognitive structures as foci of analysis. The framework represents an interactive IR process through the players (objects), their cognitive structures and relationships.

Some of the latter take place entirely at the *linguistic level* (e.g. between information objects and IR system setting), some others happen at the *cognitive level* when humans are involved.

Social and Organizational Environments. The environments are not defined in detail as concepts or variables so they remain at a pre-theoretical level of understanding. As belonging to the environments, Figure 4 suggests domains, goals, models, tasks and preferences. Domains are defined through examples, e.g., Computer Science. Goals are not exemplified but may be understood as referring to goals of activities in a domain. Models are the environment's models of the other players (Ingwersen, 1992). Work tasks also lack a rigorous definition but may be of varying complexity, more or less structured, and stable or dynamic during information access – which suggests a classification.

Individual Users. The focus is on the individuals' current cognitive states, which are variable states related to perceived work tasks, problem states, and information needs, among others. The framework is based on some evidence on the effect of work task on search process and relevance judgments (other than topical). Based on the users' knowledge levels, they may be classified into domain experts / non-experts and retrieval experts / non-experts. Their perceived work tasks may be of varying complexity.

Information needs have four types: they may be well vs. ill-defined and stable vs. variable, and the framework suggests related information access behavior, e.g., browsing in the case of ill-defined variable needs. Lack of domain knowledge suggests tasks perceived as complex and causing ill-defined needs. Information needs may be described at three levels: requests, problem statements, and work task descriptions. Poly-representation of needs at all levels is expected to lead to improved retrieval.

Information objects. Information objects are authors' texts, including titles, captions, and cited works, which are representations of authors' cognitive structures, intended to be communicated. Information objects may also be represented in various ways as different types of semantic entities (e.g., through titles, keywords, text or received citations), which may have cognitively different origins and lead to overlaps that can be systematically utilized when applied in retrieval. Information objects are of varying types, e.g., journal articles, and may thus be classified by type.

Interfaces, Requests and Queries. Interfaces support query (re-) formulation. They provide several functionalities, such as the Request Model Builder and User Model Builder functions in the Mediator Model (Ingwersen, 1992), explicitly proposed as foci of analysis. Request types reflect the four information need types as do the queries. There are other request types, mentioned in the subsection on the basic Lab IR framework, that are not explicit in the cognitive framework.

IR System Settings. The system setting consists of techniques for information object and request representations and their matching. An IR system designer's cognitive structures are embodied in specific document / query representations and matching algorithms. It is believed that it would be perfect for the system to support poly-representation.

Excluded Objects and Processes. The Cognitive IR framework subsumes the Lab IR framework (at the left side in Figure 4). The search results are not made explicit, neither evaluation, interaction is represented by arrows. However, the framework is not meant to be just an IR evaluation framework. Whether something else is missing depends on how much is seen as unexplicated within the objects presented. In his later work, Ingwersen seems to explicate many features of the framework (Järvelin & Ingwersen, 2004; Ingwersen & Järvelin, 2005).

The Model(s)

The cognitive IR approach in Ingwersen (1996) is not specific about its model(s) in the sense we use the term. There is no specific model on the Cognitive IR framework or its components. Much is open to *ad hoc* specification in individual studies. This is understandable because, on the one hand, human actors are involved, and in view of research volume, on the other. Nevertheless, if there were a specific model that would increase research volume.

The models in the Lab IR case specify that documents (requests) are practically sets of indexing features and these are represented in a specified manipulatable mathematical way. Concepts of the Cognitive IR framework, like 'social environment', 'domain', 'work task', or 'cognitive state' are obviously important in IR processes, but what are they as 'things' (or study objects) and how to 'measure' them through some variables? The simplest way is to use a nominal scale variable that provides a classification, like the variable *Domain* with possible values {'Computer Science', 'Philosophy', …}. This facilitates grouping of findings based on other variables into qualitatively different groups, which may exhibit interesting differences. However, such domains contain lots of hidden variables – observed differences cannot be explained in any deep sense by the domain *labels.* Domains are much more – but which variables does one need? There is no agreement and this makes the difficulty of modeling understandable. While observed differences in terms of information access attributes between domains-as-labels do not explain why the differences occur, such observations may still lead to asking whether the reason could be differences in discourse, vocabulary, document genres, work task types, etc. – Such problematizations may lead to more precise explanatory studies.

For some other concepts, like ‘information need’, ‘work task description’, or ‘semantic entities’, modeling may be easier. ‘Information needs’ may be classified by clarity and stability, ‘work task descriptions’ may be seen as text like requests and used alike in retrieval, and ‘semantic entities’ may be seen like documents (sources of features), but just smaller.

Laws, Theories and Hypotheses

The goals of the Cognitive IR approach “are to improve the intellectual access to information sources and, simultaneously, to provide the IR system with an enriched contextual platform that can support the user's information seeking” (Ingwersen, 1996). Therefore the goals are the same as in the Lab IR approach. Obviously, the Cognitive IR approach also seeks to explain the variation of recall and precision, but the explanatory approach is different. The focus is not only on document / request representation and matching but on information need and ‘user’ representations as well as interface functionalities. Moreover, Cognitive IR may seek to explain other phenomena than search effectiveness, e.g., the information access process.

An example of Lab IR type of explanation, while employing a different explanatory factor, relates to the principle of poly-representation, which suggests explicit representation of the ‘user's’ information need, problem state and work task – leading to different query representations. Moreover, it suggests the representation of information objects (their components) of different cognitive origin and a corresponding variety of IR techniques in order to utilize overlaps of retrieval results. The hypothesis is that objects within several overlaps are more likely to be relevant than the ones outside.

An example of non-Lab IR type of explanation relates to the classification of information needs (see above). The framework suggests that information access behavior differs by the need class. This is likely to have consequences on desirable interface functionalities, IR techniques and search effectiveness.

It might not be fair to say that the Cognitive IR approach has established well-tested *theories* of IR in the strict sense. However, at least there are hypotheses, which have received some empirical support.

Study Types, Designs and Contributions

Study Types and Contributions. The study types within the Cognitive IR approach are of all major types. There are *theoretical* studies, e.g., on relevance concepts (Cosijn & Ingwersen 2000) proposing several relevance concepts of different levels and relationships. There are *empirical* field studies such as (Cosijn, 2003), assessing the role of various types of relevance in information access – and arguing that novel relevance concepts are not just theoretical but also empirically measurable concepts. Larsen (2004) made an experiment on the effectiveness of various document sub-entities, including citations, and consequent document overlaps in the retrieval of scientific full text documents. The overlaps were shown beneficial to relevance. Larsen's study (2004) is *constructive* as well: it develops access methods employing citation networks using poly-representative overlaps both in Boolean and best-match retrieval settings. Finally, there are *methodological* studies mainly focusing on the experimental interactive IR evaluation methodology (Borlund & Ingwersen, 1998; Borlund, 2000). The latter proposes work task descriptions and an evaluation package for Interactive Information Retrieval (IIR) including novel effectiveness measures and guidelines for setting up experiments.

Study Designs. Despite of the contribution types mentioned, there is no off-the-shelf methodology for study design. The experimental laboratory studies easily follow the Lab IR study designs. The theoretical studies need to obtain their background rather from Sociology, Psychology and Philosophy than Mathematics, Statistics or Computer Science. Next we focus on the (non-laboratory type) empirical study design issues.

The empirical Cognitive IR study design is schematized in Figure 5. The nodes with darker fill, *focus areas,* are given with variables, e.g., DK_u for user's domain knowledge. Some of the nodes with lighter fill, *fringe areas,* are also given with variables. Compared to Figure 4, interaction and search results are given explicit boxes with light connecting arrows in Figure 5.

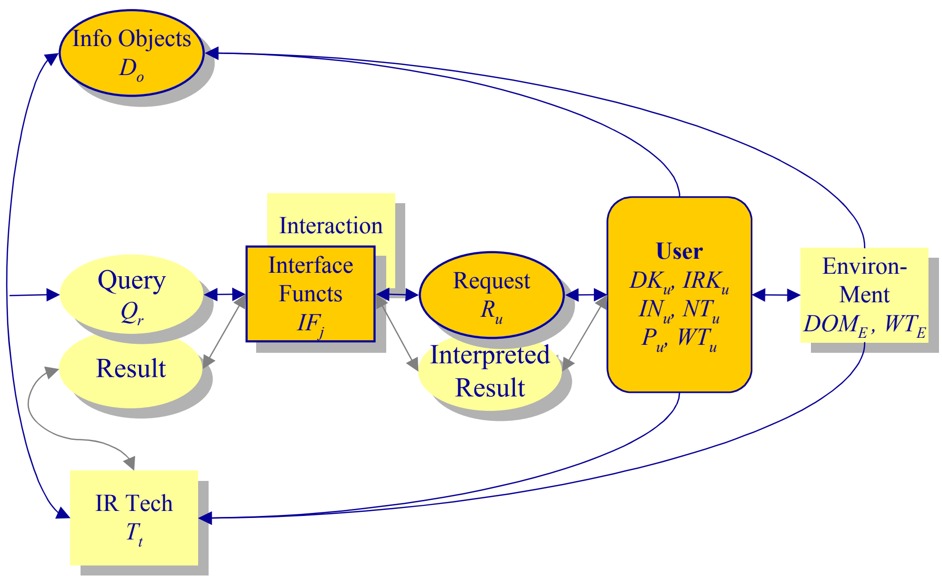

Figure 5. Cognitive IR study design (modified from Ingwersen, 1996)

The suggested foci of elaboration are related to *users*, i.e. the variables domain knowledge DK_u, IR knowledge IRK_u, information needs IN_u and need types NT_u, problem state P_u and work task descriptions WT_u; the *request* R_u reflecting the need; *interface* functionalities IF_j; and *information object* representations D_o (varying semantic entity types). We call them here proto-variables because (a) they obviously are relevant while (b) being open for definition. These variables are perhaps not an exhaustive list but nevertheless show the foci of analysis.

The fringe areas include the IR *techniques* T_t; *queries* Q_r; socio-organizational *environments* with variables domain DOM_E and work tasks WT_E; query *results*, and *interaction*. The Cognitive IR framework recognizes that there are various IR techniques (e.g. Boolean, probabilistic) that may serve different kinds of information needs differently. Likewise there are different kinds of queries (e.g. content or metadata queries) related to the requests. The environment may be described by its domain and dominant work tasks but these have so far had a minor role. Query results certainly are of interest in the framework as novel effectiveness measures (Borlund & Ingwersen, 1998) have been proposed. Interaction is a central phenomenon in the Cognitive framework, represented only by the horizontal arrows in the original framework. Therefore it remains unclear through which variables one should analyze the interaction itself. One may suggest duration, number of cycles, understandability and amount of feedback, success in terms of relevant information found, etc.

As above, one may investigate the interaction of all combinations of the variables the framework offers. In contrast to Lab IR, this framework is not geared toward explaining only the variation of a nearly fixed pair of variables. For example, one may explain the information need types by the domain and task knowledge variation. Or the variation of IR interaction (the process – however measured) by the variation of IR knowledge in natural IR environments. Of course one may trace the effects of enhanced information need representations (by work task descriptions), or the effects of enhanced document representations and overlaps, on recall and precision. There is no unique way of selecting the (in)dependent variables. Several dependent, controlled and independent variables lack standard operationalizations.

Contribution Revisited. Theoretical growth in Cognitive IR research provided the four types introduced in the section on conceptual tools above. There is *theory expansion,* e.g., due to enrichment by the concepts 'information need type' or 'work task description'. There is *greater analytical power* through systematically connecting the proposed concepts to other concepts. Moreover, there is *improved empirical support* through findings showing the explanatory power of some of the new concepts. While there are reservations regarding the definition and operationalization of variables, the Cognitive IR framework has produced many *novel research questions.*

DISCUSSION

This article has discussed two significant approaches to IR, the Lab IR approach and Ingwersen's Cognitive IR approach. Other approaches to IR may be analyzed in a similar fashion. Our research questions concerned the frameworks, models, hypotheses, laws and theories, and study designs and contributions of either approach.

Frameworks. We found that the Lab IR framework serves both modeling the IR phenomenon *and* evaluation, and focuses on IR techniques – techniques for document and request representation and their matching reflecting its disciplinary background, Computer Science. On the other hand, the *Cognitive IR framework* seeks to subsume the Lab IR framework and expands it by concepts representing interaction, users and their socio-organizational environments. It focuses on user aspects like the user's knowledge, information need types, etc.

Models. The Lab IR framework has several strong *formal* retrieval models, which specify how to represent documents and requests and how to compare these. One may choose any among them for one's study and still be able to share findings with others employing a different model. The Cognitive IR framework has no explicit model that would specify how the relevant objects and relationships must be represented for analysis. Much is left to *ad hoc* specification in individual studies. This is understandable, since several important categories are dynamic, complex and difficult to measure.

Hypotheses, Laws, and Theories. Within the Lab IR approach, hypotheses, laws, and theories are about the explanation of the variation of IR effectiveness. The explaining factors are the use or non-use of IR techniques. The approach invites one to construct theories that relate the components of retrieval models to retrieval effectiveness, that is, the use/non-use of techniques for representation and matching to recall/precision. They do not facilitate, but neither deny, exploration with various types of document collections, request types, or relevance assessment. There is an overwhelming body of research and findings, but little theory to support hypothesis formulation. *Physics* as a science is hypothetico-deductive; the non-existence of predicted particles or energy or their unexpected properties might enforce major changes in the theory of Physics. In IR it is difficult to formulate such critical hypotheses. Even if this is attempted and the hypothesis is not supported by experiments, the consequences are minor: there is no similar theoretical structure to change; the technique just failed. In lack of a coherent theoretical structure there hardly are critical hypotheses and thus the refutation of hypotheses does not matter – unlike in Physics. The only consequent problem may be the difficulty of publishing the negative result.

Within Ingwersen's Cognitive IR approach, hypotheses, laws, and theories can be about the explanation of the variation of several dependent variables. Retrieval effectiveness in terms of recall and precision is included, but the dependent variables may include variables representing information need types, information access process, or relevance assessments. Compared to the Lab IR approach by volume, the Cognitive IR approach has produced much less findings. Individual studies have produced findings on important relationships, e.g., between document polyrepresentation, query formulation, and recall/precision (Larsen, 2004), or information need representation and recall/precision (Borlund, 2000; Kelly, Dollu, & Fu, 2005), and on the existence of postulated phenomena, e.g., higher-order than topical relevancies (Cosijn, 2003). However, as in the case of Lab IR, there hardly is a coherent theoretical structure at the moment, which one could test through a critical hypothesis. At the stage of development (1996), the approach is at a proto-theoretical level. There are too many issues open for discussion instead of rigorous definitions. For example, the (possible) finding that the variable domain (e.g. one of "law", "sociology", "CS", etc.) affects some other variable in information access, say, the stability of request formulations, only provides a weak explanation because it does not detail out the mechanism of effect. Nevertheless such a finding invites one to find the mechanism, possibly through novel study designs.

Study Designs and Contributions. Within the Lab IR approach, the archetypal study design is the controlled IR experiment. In an experiment, one typically selects some variables representing IR techniques as independent variables and the query result as an intermediate dependent variable. By applying some evaluation procedure within a test collection (controlled variables), based on binary, stable topical relevance assessments and some metric, IR effectiveness becomes the final dependent variable. Strong standardization of the designs facilitates sharing and comparison of results – and make it easier to produce the next study. The limitations of the Lab approach are related to the generalizability of the findings to different kinds document collections, request types, or searcher types over various contexts of IR systems use. The control provided by test collections in their lack of variety run the risk of becoming straitjackets of research. While the variety has recently improved, using test collections without real users in context may lead to difficulties in applying the results properly in operational systems and also to suboptimal systems (Ellis, 1996). Cognitive IR challenges

Lab IR here. For example, Turpin and Scholer (2006) found out that user search task performance is hardly affected even if the search engine performance varies from 0.55 to 0.95 MAP.

In the Cognitive IR approach, there is no archetypal study design. Both field studies and lab studies are possible. Field studies need not be explanatory – they may also be exploratory. The framework is not geared only toward explaining the variation of retrieval effectiveness. There is no single way of selecting the independent, controlled or dependent variables. However, several dependent, controlled and independent variables lack standard operationalizations. Cooperation across studies and accumulation of knowledge are thus more difficult. Cognitive IR needs to be convincing about the generalizability of findings of user-oriented studies. The proponents of Lab IR probably would welcome hard-and-fast findings on users but are likely to stay off from infinite context-dependency and variability.

In theoretical study designs the background and concepts differ – formal sciences in the case of Lab IR vs. Social Sciences in the case of Cognitive IR. In methodological studies there also is a difference, since the Cognitive IR approach needs to control for human variation and thus the designs become quite different from standard Lab IR experiments, which (understandably) seek to exclude human variation.

The contributions of the Lab IR approach are novel IR techniques and an understanding on how they affect retrieval effectiveness. The contributions of Cognitive IR are findings on information access processes and effectiveness in different situations determined by user, task and need characteristics. These may lead to development of interface functionalities or retrieval techniques but may also remain as knowledge without immediate application. Shared interest in developing information access and explaining retrieval effectiveness is the connection between the approaches. Therefore the two communities seem to need to learn from each other.

Even when the Lab IR approach and the Cognitive approach explain the same phenomenon, retrieval effectiveness, however measured, the explanations tend to be very different – either through the retrieval techniques or through, say, searcher knowledge – and the community does not really know what percentage of the variation is explained by which *explanans.* Therefore there neither is a rational ground for properly focusing research efforts.

Even if Lab IR would choose to be science / technology about IR techniques and their effectiveness, it should react to the challenge on its generalizability. If one suspects the generalizability of the findings to all conceivable contexts of IR systems application, one might inform possible funding bodies that, contrary to the beliefs so far, three quarters of the IR terrain remains unmapped and thus much more funding is needed.[12] On the other hand, IR may be seen as science / technology about augmenting human task performance through improved access to information in documents. In this case there is a vast terrain to explore – and Cognitive IR is one step to that direction.

There is a gap between Lab IR and Cognitive IR. However, often progress may be attained at an intersection (or friction) point between two approaches or disciplines. And indeed, there is recent progress in interactive IR and Web IR, which seems to circumvent the gap, retain enough of the Lab IR approach to count as acceptable in Computer Science and still use concepts of the Cognitive IR approach. For example, Bell and Ruthven (2004), and White, Ruth-

[12] One *motivation* for the present exercise is a concern for IR seen as being in crisis, which was voiced in public recently at a strategic workshop on IR (SWIRL – Strategic Workshop on Information Retrieval in Lorne, Lorne, Australia, December 2004). Another concern expressed at SWIRL was that funding bodies hold IR as a solved problem and thus obtaining funding is difficult.

ven and Jose (2005) employ the concept of task complexity in the study of implicit relevance feedback in Web searching. Also link analysis and the analysis of click-through data (Joachims, 2002) can be seen as ways of analyzing socio-cognitive relevance (Cosijn & Ingwersen, 2000; Saracevic, 1996), which does not have direct indications in document texts.

Significantly changing or reorienting a research area requires efforts at all levels – in the framework, in the model, and in the research designs. Obviously, the conceptual framework needs to map a relevant part of reality as the object of study – what is seen out there sets the limits of findings. The models of the reformed approach also need to represent and relate the phenomena of interest in a way that supports study. However, the above revisions remain as lip service unless they are carried down to research designs that seek to examine the relationships of carefully defined variables. When these changes have taken place, changes in hypotheses, laws, theories and contributions will follow.

CONCLUSION

We have analyzed the Lab IR and Ingwersen's Cognitive IR approach for aspects concerned with their frameworks, models, hypotheses, laws and theories, study designs and contributions. The approaches guide IR research toward different avenues. Current Lab IR is a science / technology about IR techniques and their effectiveness. This is apparent in the Lab IR framework, which focuses on static test collections, document and request representation, and matching of the latter. Likewise, the Lab IR models are specific about modeling document and request representation, and matching. Therefore Lab IR, as a branch of technology, develops IR techniques and assesses their effectiveness. As a science, it explains the variation of recall and precision by the use of various techniques as independent variables.

Ingwersen's Cognitive IR is a (social) science about IR interaction with much focus on searchers, their cognitive aspects, and situations as factors in information access. This is apparent in the Cognitive framework, which covers a broad scope of phenomena, adding users and their environments to the Lab framework. This approach however lacks well-defined models and leaves much of conceptualization and operationalization to be specified in individual studies. As a (social) science, it may explain the variation of a number of phenomena, i.e., select the dependent, independent and controlled variables in several ways.

Significantly changing or reorienting a research area requires efforts at all levels – in the framework, in the model, and in the research designs. When these have changed, changes in hypotheses, laws, theories and contributions follow and the reformed discipline may serve individuals and the society in a new way. IR as a discipline now has great opportunities for reorientation.

Acknowledgement. This research was supported by the Academy of Finland Grant Nos. 1209960 and 204978.

REFERENCES

Allen, B. L. (1991). Cognitive research in information science: Implications for design. In: M. E. Williams (Ed.), Annual Review of Information Science and Technology: Vol. 26. (pp. 3-37). Medford, NJ: Learned Information.

Belkin, N. J. (1984). Cognitive models and information transfer. Social Science Information Studies, 4(2/3), 111-129.

Belkin, N. J. (1990). The cognitive viewpoint in information science. Journal of Information Science: Principles and Practice, 16(1), 11-15.

Belkin, N. J., Oddy, R. N., & Brooks, H. M., (1982). Ask for information retrieval: Part 1. Journal of Documentation, 38(2), 61-71

Belkin, N. J., & al. (1987). Distributed expert-based information systems: An interdisciplinary approach. Information Processing & Management, 23(5), 395-409.

Bell, D. J., & Ruthven, I. (2004). Searchers' assessments of task complexity for web searching. In: S. McDonald & J. Tait, (Eds.), Proceedings of the 26th European Conference on Information Retrieval (pp. 57-71). Heidelberg: Springer.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation, 56(1), 71-90.

Borlund, P., & Ingwersen, P. (1998). Measures of relative relevance and ranked half-life: Performance indicators for interactive IR. In: Croft, W.B. et al. (Eds.), Proceedings of the 21st ACM SIGIR Conference (pp. 324-331). New York, NY: ACM.

Buckley, C., & Salton, G. (1995). Optimization of relevance feedback weights. In: Fox, E. A., Ingwersen, P., & Fidel, R. (Eds.), Proceedings of the 18th ACM SIGIR Conference (pp. 351-357). New York, NY: ACM.

Buckley, C. & Voorhees, E. (2004). Retrieval evaluation with incomplete information. In: Jarvelin, K., Allan, J., Bruza, P. & Sanderson, M. (Eds.), SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 25-32). New York, NY: ACM.

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

Cohen, M., & Nagel, E. (1936). An introduction to logic and scientific method. London, UK: Routledge and Kegan Paul.

Cosijn, E. (2003). Relevance judgements in information retrieval. Pretoria, RSA: University of Pretoria, Dept. of Information Science, Doctoral dissertation.

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

Denning, P. J. (1997). Computer science: the discipline. In: Ralston, A. & Hemmendinger, D. (Eds), Encyclopedia of computer science, 4th Ed. Retrieved January 16, 2005 from www.idi.ntnu.no/ emner/dt8108/denning.pdf

Dervin, B. (1999). On studying information seeking methodologically: the implications of connecting metatheory to method. Information Processing & Management, 35, 727-750.

Ellis, D. (1996). Progress and problems in information retrieval. London, UK: Library Association Publishing.

Engelbart, D. (1962). Augmenting human intellect: A conceptual framework. Menlo Park, CA: Stanford Research Institute.

Hersh, W., & Over, P. (2000). TREC-9 Interactive Track Report. In: E. M. Voorhees, & D. K. Harman, (Eds.), Proceedings of the Ninth Text REtrieval Conference (TREC-9) (pp. 41-50). Retrieved October, 24, 2005 from http://trec.nist.gov/pubs/trec9/papers/t9irep.pdf

Hjörland, B. (2005). Empiricism, rationalism and positivism in library and information science. Journal of Documentation, 61(1), 130-155.

Hull, D. (1993). Using statistical testing in the evaluation of retrieval experiments. In: R. Korfhage, E. M. Rasmussen, & P. Willett (Eds.), Proceedings of the 16th Annual International ACM SIGIR Conference (pp. 349-338). New York, NY: ACM.

Ingwersen, P. (1992). Information retrieval interaction. London: Taylor Graham.

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

Ingwersen, P., & Järvelin, K. (2005). The turn: Integration of information seeking and retrieval in context. Heidelberg: Springer.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In: O. R. Zaïane et al. (Eds.), Proceedings of the 8th ACM Conference on Knowledge Discovery and Data Mining (KDD) (pp. 133-142). New York, NY: ACM.

Järvelin, K., & Ingwersen, P. (2004). Information seeking research needs extension toward tasks and technology. Information Research, 10(1), paper 212. Retrieved January 16, 2005 from http:// InformationR.net/ir/10-1/paper212.html

Järvelin, K., & Wilson, T. D. (2003). On conceptual models for information seeking and retrieval research. Information Research, 9(1), paper 163. Retrieved January 16, 2005 from http://InformationR.net/ir/9-1/paper163.html

Kekäläinen, J., & Järvelin, K. (2002). Evaluating information retrieval systems under the challenges of interaction and multi-dimensional dynamic relevance. In: H. Bruce, R. Fidel, P. Ingwersen, & P. Vakkari (Eds.), Emerging Frameworks and Method: Proceedings of the 4th International Conference on Conceptions of Library and Information Science (CoLIS 4) (pp. 253-270). Greenwood Village, CO: Libraries Unlimited.

Kelly, D., Dollu, V. D., & Fu, X. (2005). The loquacious user: a document-independent source of terms for query expansion. In: R. Baeza-Yates et al. (Eds.), Proceedings of the 28th Annual International ACM SIGIR Conference (pp. 457-464). New York, NY: ACM.

Kuhlthau, C. C. (1993). Seeking meaning. Norwood, NJ: Ablex.

Larsen, B. (2004). References and citations in automatic indexing and retrieval systems: experiments with the Boomerang Effect. Copenhagen, DK: The Royal School of Library and Information Science, Doctoral dissertation.

Pejtersen, A. M., & Fidel, R. (1999). A framework for work-centred evaluation and design: a case study of IR on the Web. Glasgow, UK: University of Glasgow, Department of Computing Science, Working paper for MIRA Workshop TR-1999-35.

Robertson, S. E. (2000). On theoretical argument in information retrieval. SIGIR Forum, 34(1), 1-10.

Salton, G., & McGill, J. M. (1983). Introduction to modern information retrieval. New York, NY: McGraw-Hill.

Sanderson, M., & Zobel, J. (2005). Information retrieval system evaluation: effort, sensitivity, and reliability. In: R. Baeza-Yates et al. (Eds.), Proceedings of the 28th Annual International ACM SIGIR Conference (pp. 162-169). New York, NY: ACM.

Saracevic, T. (1992). Information Science: Origin, evolution and relations. In: P. Vakkari, & B. Cronin (Eds.), Conceptions of library and Information Science: Historical, Empirical and Theoretical Perspectives (CoLIS 1) (pp. 5-27). London: Taylor Graham.

Saracevic, T. (1996). Relevance reconsidered '96. In: P. Ingwersen, & N. O. Pors (Eds.), Information Science: Integration in perspective, Proceedings of the 2nd International Conference on Conceptions of Library and Information Science (CoLIS 2) (pp. 201-218). Copenhagen, Denmark: Royal School of Librarianship.

Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. II. Users, questions, and effectiveness. Journal of the American Society for Information Science, 39(3), 177-196.

Sormunen, E. (2002). Liberal relevance criteria of TREC – Counting on negligible documents? In: M. Beaulieu, R. Baeza-Yates, S.-H. Myaeng, & K. Järvelin (Eds.), Proceedings of the 25th Annual International ACM SIGIR Conference (pp. 320-330). New York, NY: ACM.

Tague-Sutcliffe, J. (1992). The pragmatics of information retrieval experimentation, revisited. Information Processing & Management, 28(4), 467-490.

Turpin, A. & Scholer, F. (2006). User performance versus precision measures for simple search tasks. In: S. Dumais, E. Efthimiadis, D. Hawkins & K. Jarvelin (Eds.), SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and Development in Information Retrieval (pp. 11-18). New York, NY: ACM.

Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalization of a longitudinal study. Journal of Documentation, 57(1), 44-60.

Voorhees, E. (2001). Evaluation by highly relevant documents. In: Croft, W. B. et al. (Eds.) Proceedings of the 24th ACM SIGIR Conference (pp. 74-82). New York, NY: ACM.

Wilson, T. D. (1999). Models in information behavior research. Journal of Documentation, 55(3), 249-270.

Wagner, D., Berger, J., & Zeldith, M. (1992), A working strategy for constructing theories. In: G. Ritzer (Ed.), Metatheorizing (pp. 107-123). Sage.

Wang, P., & 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(2), 115-133.

White, R. W., Ruthven, I. & Jose, J. M. (2005). A study of factors affecting the utility of implicit relevance feedback. In: R. Baeza-Yates et al. (Eds.), Proceedings of the 28th annual international ACM SIGIR Conference (pp. 35-42). New York, NY: ACM.

Witten, I. H., Moffat, A., & Bell, T. C. (1999). Managing gigabytes: Compressing and Indexing documents and images, 2nd ed. San Francisco, CA: Morgan Kaufmann.