UNIVERSITY OF COPENHAGEN

A context-driven integrated framework for research on interactive IR

Ingwersen, Peter

Published in:
Tushu Qingbao Zhishi

Publication date:
2008

Document version
Også kaldet Forlagets PDF

A Context-Driven Integrated Framework for Research on Interactive IR[1]

Peter Ingwersen

Royal School of Library and Information Science
Birketinget 6, DK Copenhagen 2300 S – Denmark
pi@db.dk

Abstract

This paper discusses the Integrated Research Framework for Information Seeking and Retrieval (ISR) originating from (Ingwersen & Järvelin, 2005) by comparing it to the Laboratory Research Framework for IR and two nested models of contexts involved in ISR: that of Kekäläinen & Järvelin (2002), based on work task activities, and the model by Ingwersen (2007) focusing on contexts to information objects. In addition the relevance model by Cosijn (2004), also of nested nature, is discussed, leading forward to the nine dimensions of research variables, potentially influencing ISR processes and evaluation.

1. Introduction

The Integrated Research Framework for Information Seeking and Retrieval (ISR), proposed by Ingwersen & Järvelin (2005), regards information retrieval (IR) as embedded in information seeking behavior. The latter is seen as contextual to the former and itself in context of the job-related or daily-life work task situations and domains. A research framework is understood to support researchers generating research designs. Research design consists of (a) the research problem, (b) data collection, and (c) research methods. The Integrated Research Framework and the research designs aim at (i) IR system design and evaluation as final goals or (ii) understanding human (searcher) behavior in the context of tasks, domain and information systems. It seeks to emphasize the central components of investigation and their features, the *research variables*, in interactive IR, i.e., the algorithmic part of IR systems, interfaces and information objects, as well the interaction process itself and searchers in their context – see Figure 1. In short, the Integrated Research Framework aims at providing the participants with a research methodology for investigating central facets and their relationships of IR interaction.

Research frameworks contain (tacit) shared assumptions on ontological, conceptual, factual, epistemological, and methodological issues. A Research Model is a precise (often formal) representation of objects and relationships (or processes) within a framework. Modeling may also in principle encompass human actors and organizations. In mainstream IR research there exists only one robust research framework: The Laboratory Research Framework for IR, Figure 2, but many different mathematical/algorithmic IR models compete, like the probabilistic model, language models or vector-space models.

In more user-oriented IR research or in information seeking studies several different research frameworks (and conceptual models) compete. They are commonly concentrating on different aspects of information interaction, including IR interaction. Where the Laboratory Framework's strength derives from the lack of a participating seeking actor(!), the user-centered frameworks and

---

[1]This paper was originally presented as keynote at the LIDA Conference, June 25-30, 2008, Dubrovnik.

models indeed contain such actors but scarcely model the algorithmic part of the frameworks. The Laboratory IR Framework is thus quite simplistic, but robust and demonstrates control over a few well-known variables; the latter frameworks are complex and full of more or less vaguely defined variables that are difficult to control.

The present contribution seeks to demonstrate how Ingwersen & Järvelin's Integrated Research Framework for ISR in context combines the two kinds of research frameworks and models into a solid unified entity that encompasses mainstream IR, interactive IR and seeking research. The main focus is on the set of variables that are involved in such studies and how much context that is required for investigative purposes. The ensuing section discusses briefly the general model of the Integrated Research Framework and the mainstream Laboratory IR Framework as part of the former. This is followed by a discussion of the contextual properties and borderlines for the Integrated Research Framework. Section 4 demonstrates the Integrated Research Framework's dimensions and range of variables that must be taken into account in research on ISR. The contribution concludes with a brief discussion of the application of the framework.

2. Integrating Research Frameworks

Figure 1 displays the conceptual model of the Integrated Research Framework for ISR research (Ingwersen, 2007). It relies on a cognitive perspective of information interaction in context, i.e., it aims at observing the cognitive actors, their relationships and manifestations (or representations of cognition) in ISR. In each component of the model different actors may contribute interpretations of the world that together form the features of that component. The Seeking Actor (or team) is one such central component; another is the Information Objects generated by actors in the role of authors over time. The shape of the model indicates that not only the Socio-Organizational or Cultural environment serves as context for the seeking actor and the information systems (Objects, IT and Interface). The 'Systemic' left-hand side of the model likewise serves as context to the other components. This balance is emphasized by the essential component of *Interaction*, connecting the two major contextual elements of the model.
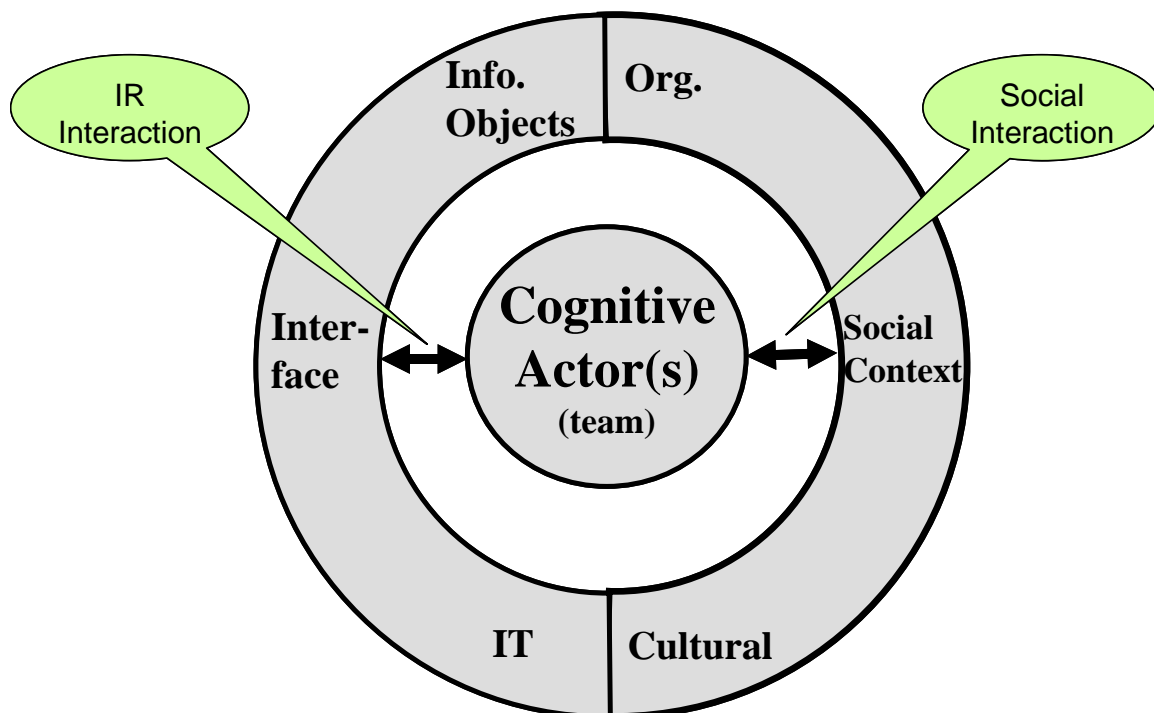
Figure 1. The Circle of Systemic/Social Contexts in Information Seeking & Interactive IR (extension of Ingwersen, 2007, p. 15).

The Laboratory IR Research Framework is depicted on Figure 2. At its core its central Laboratory IR *Model* is situated. That model consists of a set of documents and a request. Both become represented by identical algorithmic indexing principles (requests are seen as small documents), ending up in a database and a query that are matched. If required after this first run of the search engine, 'pseudo relevance feedback' may be done to start a second retrieval run by a modified query version. Pseudo relevance feedback implies automatically to feed back to the retrieval algorithm a defined set of assumingly relevant documents, for instance the top-five documents taken form the matching result (commonly ranked by algorithmic relevance (Saracevic, 1996;)). The remaining features of the Laboratory Framework refer to the IR evaluation process of assessing the retrieved documents by matching these against a recall database consisting of relevant documents pooled in relation to a set of predefined requests. These relevance assessments are commonly (e.g. in TREC) done *posteriori* to the retrieval processes by experts. The predefined requests are commonly named 'topics'. Metaphorically speaking the framework, Figure 2, is embedded within a (laboratory) cave, with the real searcher and other contextual elements settled outside the cave. The central Laboratory Model *and* the more extensive Laboratory Framework for IR research are nested within the Integrated Research Framework, the left-hand side of the model, Figure 1. Järvelin (2007) discusses the theoretical perspectives of the Laboratory Framework compared to an earlier version of the Integrated Research Framework, the cognitive model for interactive IR (Ingwersen, 1996).
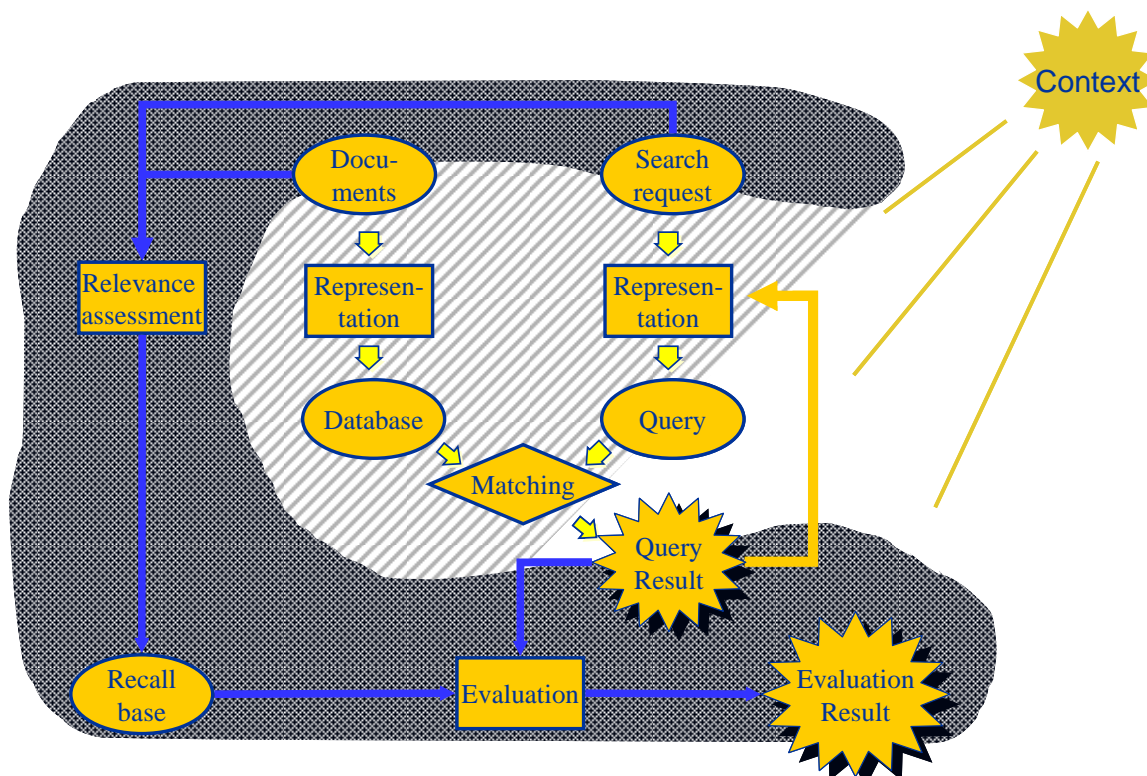

Figure 2. The Laboratory Research Framework for evaluating IR outcomes, including the Laboratory Model (From Ingwersen & Järvelin, 2007, p. 137)

Figure 3 stresses this integration in the form of a nested model (Kekäläinen & Järvelin, 2002). The Laboratory IR Model is seen in context of Information Seeking processes, for which IR only

serve as one of many means to access to information. Information Seeking is nested in the Work Task context, as perceived by the seeking actor, and further out the task processes are nested in the Socio-organizational and Cultural context.

This nested model differs from that by Wilson (1999) in that it takes into account additional contextual features including job-related or daily-life work tasks that influence search tasks, searching and interaction activities and retrieval processes. Wilson' Model focuses solely on Information Seeking in context of Information Behavior. As such, the nested model, Figure 3, is more closely associated to Saracevic Stratified Model (1996), which laid the foundation for his relevance typology and levels of interaction.

The strength of the nested model, Figure 3, is its involvement of the work task, the processes to be conducted in order to fulfil such tasks, as well as pointing to the task result. Seeking (and retrieval) tasks become thus influenced by work tasks, and the model points to the necessity of developing novel information outcome (evaluation) measures, aside from, e.g., utility measures like satisfaction and recall/precision measures, in order to meet the challenges of the work task environment.
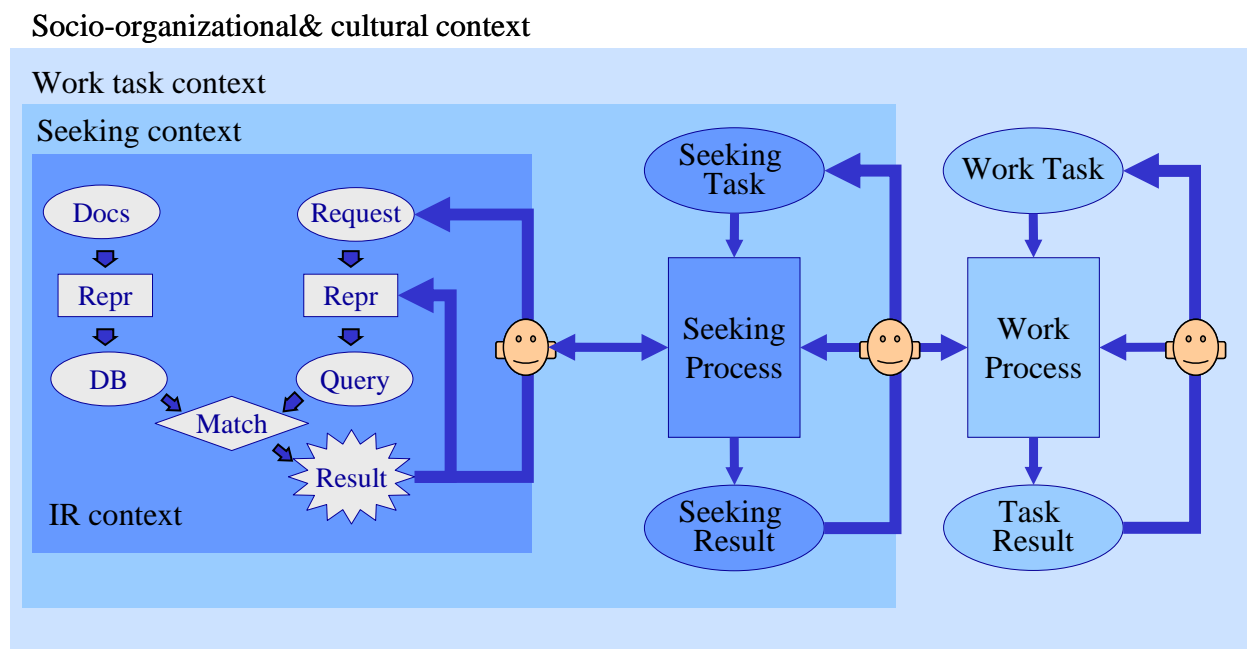

Figure 3. Nested model of IR and Information Seeking activities in context of Work tasks and Socio-organization & Cultural environments (From Kekäläinen and Järvelin, 2002, p 263).

2.1 Relevance Typology and the Integrated Research Framework

Cosijn constructed an alternative contextual model (2004) – Figure 4. The model is very detailed and highly associated with the nested model, Figure 3, the conceptual Integrated Research Framework, Figure 1, and the Saracevic Stratified Model (1996). She developed the model in order to discuss the locations of the Saracevic (1996), Cosijn & Ingwersen (2000) and Borlund (2003a) relevance categories.

In the upper left corner the Laboratory Model for IR is located as a kind of black box. It's opening toward context holds an interface (and potentially a human intermediary). Within the box 'algorithmic relevance' assessments are made by the retrieval algorithm(s) in question – resulting in

an output via the interface. The idea behind the model is that when moving *from* the contextual features towards the system (from lower right corner) searching activity and retrieval take place, including query modification loops. One may observe the social interaction taking place between the individual and his/her social context and the perception of work task from that context. For each box in the Information Seeking environment new perceptions and modifications of the search activity and contents take place. In the opposite direction, from the IR system *towards* the contextual features at the right-hand part of the model information use takes place and the relevance categories, according to Cosijn & Ingwersen (2000), come into play. This is the strength of the model.

The difference between the Saracevic (1996) and the Borlund (2003a) relevance perceptions and the Cosijn interpretation lie in the 'socio-cognitive' relevance conception (Cosijn & Ingwersen, 20000). Where 'situational relevance' is highly individual the 'socio-cognitive' type of relevance signifies that groups or networks of actors assign relevance to a particular object, like done over time in the form of scientific citations or as recommendations via extractions from search engine logs. One might argue that aside from citation analysis also data mining takes advantage of socio-cognitive relevance structures of various kinds. Both types of analyses are attempts of knowledge management to control and visualize an increasingly vast, divergent and quality-dispersed flow of information. See also the profound literature reviews by Saracevic (2007a+b).
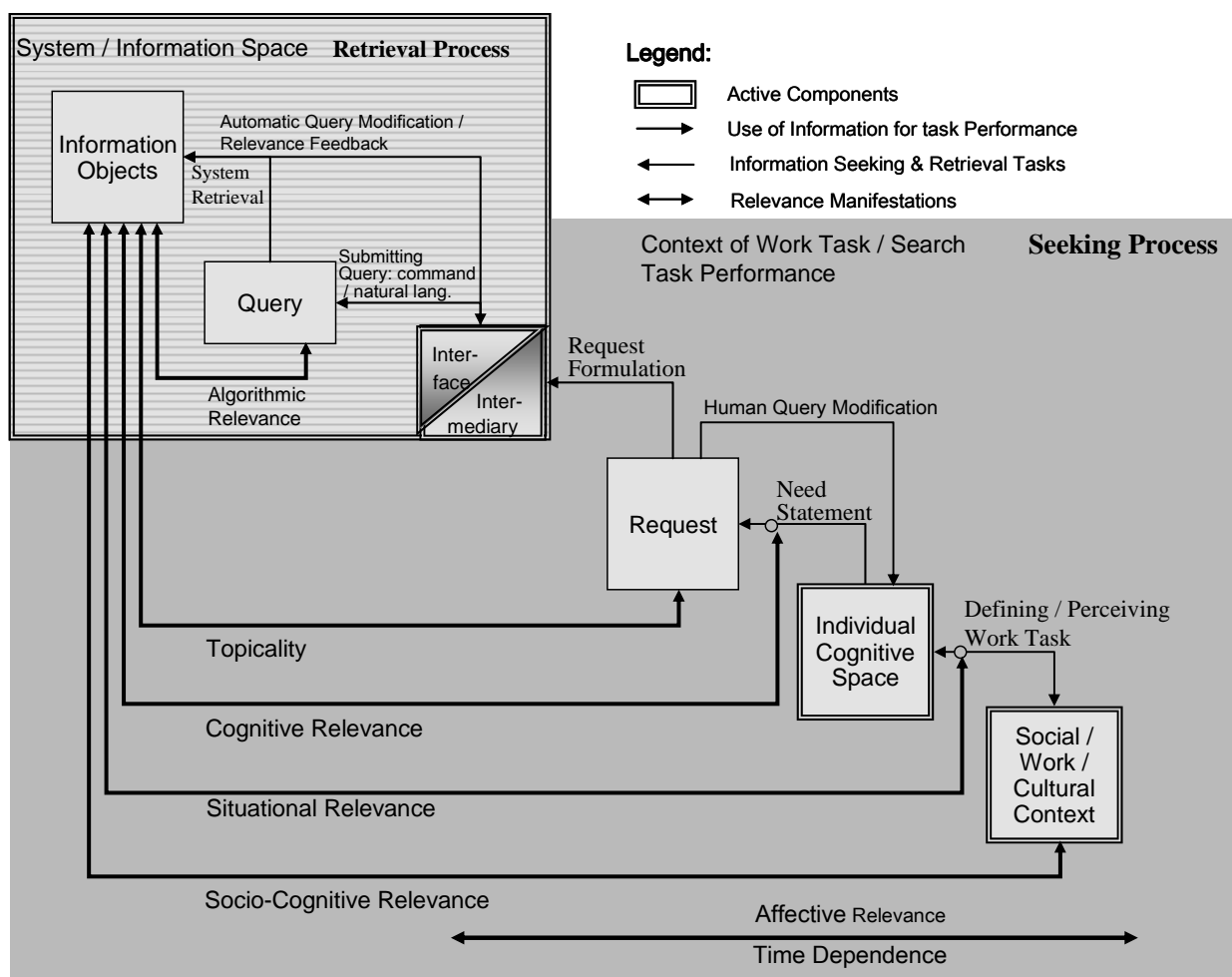

Figure 4. Contextual relevance typology and assessment model for ISR (From Cosijn, 2006, p. 22)

3. Where Does Context Begins or End?

Obviously, for each component of the models Figures 1-4, as well as for the *interaction process* itself, one may depict sets of research variables that may be included into research designs and influence the methodological approach to a given project. One problem to be addressed prior to the outlining and discussion of such variables is to define the borderline of context that may be taken into account. All the models have hitherto mainly concentrated on the IR system, the searching actor, the work task situation perceived by the actor and, to an extent, on the socio-organizational and cultural environment; but what about the broader societal environment? How far 'out' does one have to capture relevant context features? See also Kelly (2006a+b) for an in-depth research review.

Contextual features should be involved as long as they are (1) *neighboring* the component in focus of the investigation in question, see Figures 1-4, and (2) *detectable* as potential biases in scientific investigations. This implies that if the central issue of a study concerns, say, retrieval of information objects, variables from the neighboring components should be taken into account, Figure 5, i.e., at least algorithmic elements of the IR system, the processes of retrieval and interaction , including request and query types, and seeking actor characteristics. A few such variables are independent while most should, ideally, be controlled or neutralized. With 'detectable' is meant that if a range of studies in a research field only detects insignificant influence from a (set of) contextual feature, at a given point in time, then such features should not be taken into account – even though they may be influential in a theoretical/epistemological sense. However, later IT development or other circumstances may indeed make such features count in studies implying that one must include them at that time. An example of such features taken from the 'seeking actor's' context is 'display time' during IR interaction which, so far, does not seem associated with (positive or negative) relevance assessment (Kelly & Belkin, 2004).
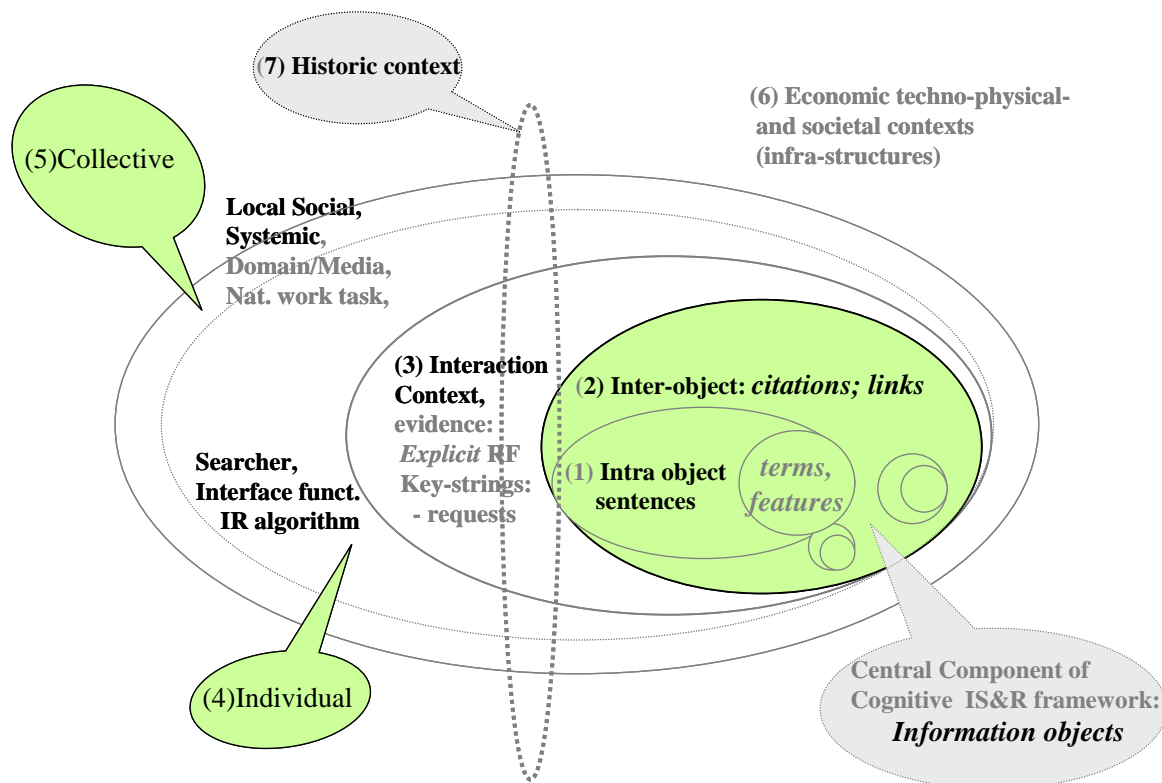

Figure 5. Nested model of context types centered on Information Objects of an IR system (Ingwersen, 2007, p. 19; extension of Ingwersen & Järvelin, 2005).

The model Figure 5 re-organizes the components of the previous models, Figures 1-4, as nested depending on which focus that is stressed in an empirical study. The particular configuration on Figure 5 centers on the 'information objects' and demonstrates which neighboring contextual elements that may be interesting to pursue in a study. In Ingwersen (2007) other configurations are demonstrated and discussed.

There are six sets of nested context categories (Ingwersen, 2007, p. 18):

1. *Intra-Object structures*: terms, phrases, image features, pixels, sentences , paragraphs, sections …
2. *Inter-Object contexts*: links, citations, document clusters defined of various features …
3. *Interaction (session)*: search/authoring process *evidence*, e.g., eye/mouse movements, Work Task descriptions, explicit Relevance Feedback, search task path; features of algorithmic IR processes, auto-indexing keys-strings; …
4. *Individual* context, dependent on Central component (subjective): actor, current Work Task perception … ;(objective): engine logic/algorithms, interface functionality, …;
5. *Collective* local context, dependent on Central component: (local) socio-org.-cultural structures/conditions: domain vocabulary, natural Work Tasks, organizational preferences, several searchers' Work Task perceptions, their socio-cognitive relevance assessments and recommendations, implicit RF behavior; (local) systemic conditions , …
6. *Techno-economic-politico-societal infrastructures* influencing (not necessarily always in a remote way) all actors, components and interactive sessions;

In addition, the *historic context* operates across this stratification, i.e., the history of *all* participating actors' experiences, forming their expectations. All IIR processes and activities are under influence of this temporal form of context.

Evidently, there exists a practical limitation in research designs concerning context when we are approaching the borderline of the collective socio-organizational & cultural context (5) towards societal and economic or global contexts (6), Figure 5. However, as pointed out by Dick when applying this model to analyze the information situation in developing countries (2005): the technical and political *infrastructure* (6) is observed as a strong determining factor for how information work (and ISR) is carried out in such countries and their institutions. On this national or regional level the techno-economic structures and the historic context play vital roles for all kinds of understanding of development. In a more narrow sense of information studies, related to ISR, research design commonly involves variables from the context components 1-4. The Integrated Research Framework for ISR also includes variables from the '*Collective local context*' (5) because, in accordance with empirical observations and thus the models Figure 1, 3-4, this context provides hard evidence of influence on the other model components. Searchers generate information needs based on perceptions of job-related or daily-life tasks. In certain knowledge and human resource heavy organizations such work and search tasks objectively exist as parts of local knowledge management systems, like in pharmaceutical and biological companies – as Good Manufactory Practice or Good Laboratory Practice and like knowledge sources. Also recordings of how local employees have acted during the work process or logs of how and what millions of local or global searchers have searched form part of this kind of context.

4. The Integrated Research Framework for ISR

A total of nine dimensions of variables constitute the Integrated Research Framework. They associate to the components of the models, Figure 1-4, and the contexts (1) through (5), Figure 5.

Table 1 displays the dimensions and their associations to the ISR components. Each dimension holds a number of variables, each potentially carrying more than one value – Tables 2-3.

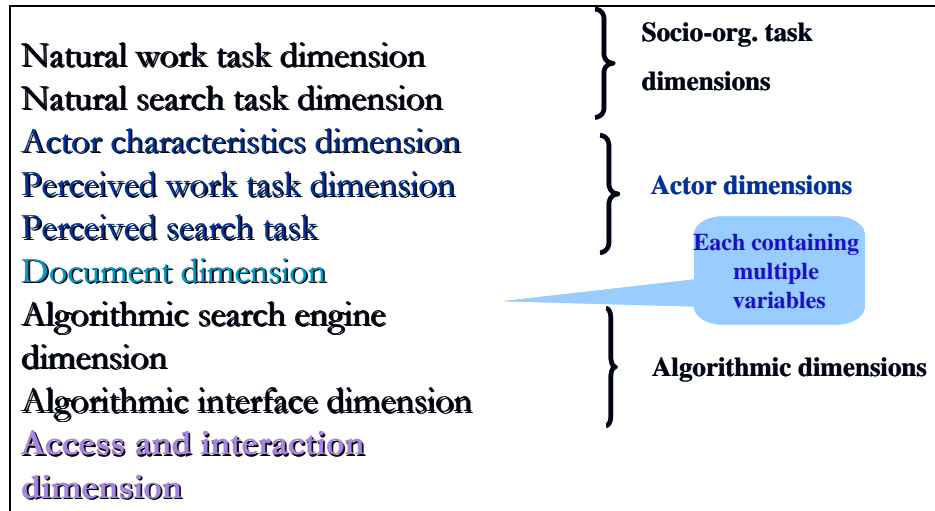Table 1. The nine dimensions of the Integrated Research Framework for ISR.

Table 2. The research design cube for nine IS&R research variables (Ingwersen & Järvelin, 2005, p. 360).

Natural Work Tasks (WT) & Org	Natural Search Tasks (ST)	Actor	Perceived Work Tasks	Perceived Search Tasks
WT Structure	ST Structure	Domain Knowledge	Perceived WT Structure	Perceived Information Need Content
WT Strategies & Practices	ST Strategies & Practices	IS&R Knowledge	Perceived WT Strategies & Practices	Perceived ST Structure/Type
WT Granularity, Size & Complexity	ST Granularity, Size & Complexity	Experience on Work Task	Perceived WT Granularity, Size & Complexity	Perceived ST Strategies & Practices
WT Dependencies	ST Dependencies	Experience on Search Task	Perceived WT Dependencies	Perceived ST Specificity & Complexity
WT Requirements	ST Requirements	Stage in Work Task Execution	Perceived WT Requirements	Perceived ST Dependencies
WT Domain & Context	ST Domain & Context	Perception of Socio-Org. Context	Perceived WT Domain & Context	Perceived ST Stability
		Sources of Difficulty		Perceived ST Domain & Context
		Motivation & Emotional State		

　　Two dimensions relate to the socio-organizational & cultural context component: the 'natural work task' set of variables *and* the 'natural search task' dimension. These two sets of variables concern the objective work and search tasks, briefly discussed above, and existing in job and daily-

life environments. Next, the actor component consists of three dimensions. Two associated with to *perceptions* of the 'natural work/search task' variables. One dimension of the actor concerns his/her/their personal characteristics, such as level of domain expertise and seeking knowledge.

The 'document dimension' outlines the known variables connected to document features, such as 'Document Type', 'genre' or 'structure'. Two 'algorithmic dimensions' of variables exist: One dealing with the search engine(s) and their algorithmic retrieval models or 'use of weights' in automatic indexing, and one dimension on interface variables.

Table 3. Research design cube continued … (Ingwersen & Järvelin, 2005, p. 361)

Document and Source	IR Engines IT Component	IR Inter-faces	Access and Interaction
Document Structure	Exact Match Models	Domain Model Attributes	Interaction Duration
Document Types	Best Match Models	System Model Features	Actors or Components
Document Genres	Degree of Doc. Structure and Content Used	User Model Features	Kind of Interaction and Access
Information Type in Document	Use of NLP to Document Indexing	System Model Adaption	Strategies and Tactics
Communication Function	Doc. Metadata Representation	User Model Building	Purpose of Human Communication
Temporal Aspects	Use of Weights in Doc. indexing	Request Model Builder	Purpose of System Communication
Document Sign Language	Degree of Req. Structure and Content Used	Retrieval Strategy	Interaction Mode
Layout and Style	Use of NLP to Request Indexing	Response Generation	Least effort Factors
Document Isness	Req. Metadata Representation	Feedback Generation	-
Document Content	Use of Weights in Requests	Mapping ST History	
Contextual Hyperlink Structure		Explanation Features	
Human Source (see Actor)		Transformation of Messages	
		Scheduler	

5. Discussion

Some of these variables have been investigated in many research scenarios, e.g. the expertise and knowledge levels of searchers. Others are scarcely or not at all studied in natural settings, such as 'granularity' of work tasks.

One may relate single or groups of dimensions to the models of (interactive) IR previously displayed. In the Laboratory Framework for IR, Figure 2, the central dimensions are '*Document and Source*' and '*IR Engines, IT Component*'. When comparing to the actual research done in mainstream IR we observe that only recently laboratory IR touches upon variables like 'Document Genres' or 'Document Isness' (metadata). Only very rarely do the different variables become cross tabulated, e.g. that one compares different document types and genres, connected to document

structure and layout tested on different retrieval models, in order to understand each model's advantages and impact on IR performance.

The 'Interface' dimension is commonly not taken into account in the Laboratory IR Framework, but one observes an increasing awareness of the influence of request types (Request Model Builder' function) and information need types on IR outcomes. In so-called IR interaction 'light' (Ingwersen & Järvelin, 2007) one incorporates contextual dimensions, Figures 1, 3-4, such as the 'Actor' and 'Perceived Work/Search Tasks' in session-based search scenarios. In a more 'ultra-light' version of IR interaction, still within the Laboratory Framework, Figure 2, a seeking actor may only carry out one run providing relevance feedback, or one simulates searcher behavior based on *a priori* knowledge of contextual features. This is owing to the learning effects on test persons beyond a single retrieval run. Several consecutive runs make indeed the research design realistic but, at the same time, make the retrieval outcome incomparable with the already established relevance assessments in the 'Recall Base, Figure 2.

IR interaction 'light' involves searchers who become responsible for the relevance assessments according to the Cosijn Model, Figure 4. Commonly, 'light' IR research designs may involve assigned topics, or better 'simulated work task situations' or cover stories, as proposed and tested by Borlund (2003b). The strength of the simulated task situations is that the research scenario maintains a certain degree of experimental control, but allows for individual freedom as to information need and request formulations by the test persons. The disadvantage is that one has to be very careful when replacing real task or information need situations by simulated ones. They have to designed in such a way that the motivation and contextual situation is felt 'natural' to the participants.

In completely naturalistic field studies of IR interaction and information seeking behavior all the nine dimensions may contribute variables to be selected as independent ones, controlled or neutralized ones or they are simply '*hidden variables*', potentially biasing the outcome. In Ingwersen & Järvelin (2005; 2007) are provided different research scenarios and designs according to a variety of research questions and exemplifying the application of the Integrated Research Framework for ISR.

References

Borlund, P., (2003a). The concept of relevance in IR. Journal of the American Society for Information Science and Technology, 54(10): 913-925.

Borlund, P. (2003b). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. Information Research, 8(3), paper no. 152 [Available at: http://informationr.net/ir/8-3/paper152.html. Cited May 13, 2003.]

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

Cosijn, E. (2006). Relevance judgments within the context of work tasks. Information Interaction in Context, IIiX, Copenhagen, Royal School of Library and Information Science, Denmark (ACM DL): 20-29.

Dick, Archie (2005). Power is information: South Africa's promotion of access to Information Act in context. In: F. Crestani & I. Ruthven (eds.), Context: Nature, Impact & Role: 5th International Conference on Conceptions of Library and Information Science – CoLIS 5. Springer :212-225

Ingwersen, P. (2007). Context in information interaction – revisited 2006. In: Bothma, T. & Kaniki, A. Prolissa 2006: Proceedings of the Fourth Biennial DISSAnet Conference. Farm Inn, Pretoria, Infuse: 13-23.

Ingwersen, P. & Järvelin, K. (2005). The Turn: Integration of Information Seeking and Retrieval in Context. Springer, Netherlands.

Ingwersen, P. & Järvelin, K. (2007). On the holistic cognitive theory for information retrieval. In: Dominich, S & Kisss, F. (eds.), Studies in Theory of Information Retrieval. Budapest: Foundation for Information Society, 2007: 135-147.

Järvelin, K. (2007). An analysis of two approaches in information retrieval: from frameworks to study designs. Journal of American Society for Information Science and Technology, 58(7): 971-986.

Kekäläinen, J. & Järvelin, K. (2002). Evaluating information retrieval systems under the challenges of interaction and multi-dimensional dynamic relevance. In: Bruce, H., Fidel, R., Ingwersen, P. & Vakkari, P. (eds.), Emerging

Frameworks and Methods: Proceedings of the 4th International Conference on Conceptions of Library and Information Science (CoLIS 4). Greenwood Village, CO, USA, Libraries Unlimited: 253-270.

Kelly, D. (2006a). Measuring Online Information Seeking Context, Part 1: Background and Method. Journal of American Society for Information Science and Technology, 57(13):1729–1739.

Kelly, D. (2006b). Measuring Online Information Seeking Context, Part 2: Findings and Discussion. Journal of American Society for Information Science and Technology, 57(14):1862–1874.

Kelly, D., & Belkin, N. J. (2004). Display time as implicit feedback: Understanding task effects. In K. Järvelin, et al. (Eds.), Proceedings of the 27th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '04). New York: ACM: 377–384.

Saracevic, T., 1996. Relevance reconsidered '96. In: Ingwersen, P. and Pors, N.O. (Eds.), Information Science: Integration in Perspective: 2nd CoLIS Conference. Royal School of LIS, Copenhagen: 201-218.

Saracevic, T. (2007a). Relevance: a review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. Journal of American Society for Information Science and Technology, 58(13): 1915–1933.

Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effect of relevance. Journal of American Society for Information Science and Technology, 58 (13): 2126–2144

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