

INFORMATION RETRIEVAL *I*NTERACTION

PETER INGWERSEN

Contents, Preface & Introduction

This book is available for download at [www.db.dk/pi/iri]



© Peter Ingwersen 1992, 2002

Peter Ingwersen

E-mail: pi@db.dk

Homepage: www.db.dk/pi

Royal School of Library and Information Science

Department of Information Studies

Birketinget 6

DK-2300 Copenhagen S

Denmark

Information Retrieval Interaction was first published in 1992 by Taylor Graham Publishing. This electronic version, published in 2002, was converted to PDF from the original manuscript with no changes apart from typographical adjustments. It has been ensured that the page numbering of the electronic version matches that of the printed version. Both versions can therefore be cited as:

Ingwersen, P. *Information Retrieval Interaction*. London: Taylor Graham, 1992. X, 246 p.

Published in 1992 by
Taylor Graham Publishing
500 Chesham House
150 Regent Street
LONDON W1R 5FA
United Kingdom

Taylor Graham Publishing
12021 Wilshire Boulevard
Suite 187
LOS ANGELES, CA 90025
USA

ISBN 0 947568 54 9

Converted to PDF in November 2002 by Rasmus Bruun

CONTENTS

Preface	V
Acknowledgements	V
Introduction	VII
CHAPTER 1	
Information Science in Context	1
1.1 The emergence of information science	1
1.2 The scope of information science	11
CHAPTER 2	
The Cognitive View and Information	15
2.1 The cognitive viewpoint versus cognitivism	19
2.2 Understanding information in information science	26
2.3 Information science seen as a cognitive science	37
2.4 Hermeneutics in IR interaction	41
2.5 Summary statements	47
CHAPTER 3	
Information Retrieval	49
3.1 Essential issues in IR	49
3.2 Major IR research approaches – an overview	57
CHAPTER 4	
The Traditional IR Research Approach	61
4.1 Classification theories	63
4.2 Indexing theory, controlled vocabulary issues	64
4.3 Natural language representation	67
4.4 IR technique developments	72
4.5 Relevance measurement techniques	78
4.6 Summary statements	80
CHAPTER 5	
The User-oriented IR Research Approach	83
5.1 The role of the intermediary and the user in IR	87
5.2 Major empirical user studies	91
5.3 Empirical investigations of user-intermediary-system interaction	99
5.4 Pre-search interviewing investigations, excluding retrieval	105
5.5 User-oriented analytic studies of IR interaction	112
5.6 Summary of analytic and empirical user-oriented studies	121
CHAPTER 6	
IR Interaction – the Cognitive Turn	123
6.1 Personal cognitive structures relevant to IR	124
6.2 Cognitive models for IR interaction	124
6.3 Design and evaluation methodologies in IR interaction	146
6.4 Summary statements	156
CHAPTER 7	
The Cognitive IR Research Approach	157
7.1 Selected intermediary designs in IR	160
7.2 ‘Intelligent’ user model building vs supportive mechanisms in IR	175
7.3 The supportive user model building approach	181
7.4 Knowledge-based adaptive IR interaction	186
7.5 Towards a contextual IR theory?	194
7.6 Summary	201

CHAPTER 8	
The Mediator Model	203
8.1 The framework of the Mediator Model	204
8.2 The use of the framework	220
CHAPTER 9	
Conclusions	223
Definitions	227
References	231
Subject Index	241

PREFACE

Books are not made to be believed, but to be subjected to inquiry.
When we consider a book, we mustn't ask ourselves what it says but what
it means...

Umberto Eco: *The Name of the Rose*, 1984, p. 316.

Information retrieval covers the problems relating to the effective storage, access, and searching of information required by individuals. Currently, information is continuing to grow exponentially, diversifying into many forms and media. In this complex retrieval labyrinth there is a definite need for increased effort aimed at tailoring IR performance to user demands.

As Umberto Eco makes the learned Brother William point out in a moment of reflection visiting the library, the fundamental problem in information retrieval is how to bridge text and its potentiality for providing information to the individual reader.

As a contribution to these continuing efforts of harmony between information and user, the objective of this publication is to exhibit and enhance the theoretical and operative requirements necessary for effective performance, in particular of intermediaries, in information retrieval interaction.

For me this book represents a turning point. It covers more than ten years of progressive research and development work, inspired and supported by colleagues and friends in an international environment.

ACKNOWLEDGEMENTS

During my work on this book I have received encouragement and criticism from many colleagues. This has been extremely advantageous in forming and cultivating my concepts and ideas.

At this moment in particular my thoughts go to the late Professor Povl Timmermann whose visionary and innovative ideas about information retrieval research originally made me initiate this interesting exploration. Throughout the years they have often been my guide. Moreover, I want to express my appreciation of the adaptive way Professor Niels Bjørn-Andersen has provided me with his constructive support, which has been extremely beneficial during this period of creativity.

Finally, but most important, I want to thank my wife Irene Wormell, for her intellectual sustenance as well as for all the necessary logistics.

Peter Ingwersen, Ph.D., FLA

Copenhagen, 1992.

[This page intentionally left blank]

VI

During recent years the author has evaluated several international projects, mainly originating from outside the core information retrieval (IR) environment. They have been concerned with the design and management of a variety of information systems ranging from knowledge-based applications to office automation configurations incorporating hypermedia and hypertext.

In common to such projects one may notice that the chosen retrieval components rather often constitute the weak elements in otherwise solid proposals. The suggested IR techniques and the indexing methods adhere to decade-old traditions, and the proposed interface functionalities may demonstrate a certain lack of functionality.

The IR field itself actually has produced sound and inspiring monographs and articles on IR theory, research and applications. However, the monographs are often not recently published, and for this reason alone difficult to obtain, or they demonstrate stand-alone approaches to IR research. Similarly, the journal articles and thematic reviews, indeed the IR research itself, seem to demonstrate a diversity of smaller communities, each one viewing IR from their own position.

In the author's view this somewhat blurred state of affairs in IR research, not really demonstrating a unifying framework for the field, is rather unfortunate. It results in a much smaller export ratio to related fields than deserved. The actual moment is ripe for a change: the technological advance has opened up for further *integration* of very different information sources and processes on a larger scale than hitherto observed in the entire information sector; also IT itself is not seen any more as *the* solution to the organisation, provision and use of information, not even in a multimedia context. The focus of attention has moved into the *qualitative* aspects of such processes. Intellectual access and use of information – structured as well as unstructured, administrative as well as textual and image-based – are the requirements asked for today.

This situation creates an optimal opportunity for IR, provided that the field overcomes its present state of theory fragmentation. Viewed as a variety of technical, intellectual and conceptual elements that actually should be fitting together, IR holds a strong potential for successful contributions to integrated systems design in future.

One of the reasons for the diversification into smaller self-contained IR research communities may be that information science as a discipline until recently has suffered from similar fragmentation problems. As an educator and developer of curriculum programs in library and information science (LIS) one is very often asked questions from colleagues, such as: 'What is information science really – what is its substance?' – or 'IR, that's simply a technique, isn't it?' Again, the fragmentation of the information landscape makes it rather cumbersome to provide simple and exhaustive answers, for instance to fellow colleagues or to researchers in other fields.

To illustrate the colourful cocktail of IR research and its findings one picks the basic ingredients, Salton and Sparck Jones, mixing them with van Rijsbergen, whereby mathematical-linguistic-logico positions in IR theory from the past and present are covered. Then, this can be mixed with some Belkin & Vickery and Ellis to provide a user-orientation in a socio-psychological context, and the cocktail is almost ready by adding pragmatic aspects from Croft or E. Fox. To give it a final touch one adds some drops of Blair (Wittgenstein) and the mandatory Winograd & Flores (Heidegger). This blend or other similar ones can be expected to be created in genuine LIS departments all over the world – not easily achieved outside our field, however.

In the attempt to compensate for the situation outlined above it is hoped that this publication will be of value to the IR community and in addition can be used in related fields, providing an integrated understanding of solutions to IR problems and contributing to the progress of the field.

Hence, the aims of the book are to establish a unifying scientific approach to IR – a synthesis based on the concept of IR interaction and the cognitive viewpoint; to present research and developments in the field of information retrieval based on a new categorisation; and to generate a consolidated framework of functional requirements for intermediary analysis and design – the Mediator Model. The introduction describes the aims and the organisation of the contents, including an outline of the original contributions to the field of IR presented in this book.

IR interaction is defined as the interactive communication processes that occur during the retrieval of information by involving *all* the major participants in IR, i.e. the user, the intermediary, and the IR system – the latter consisting of potential information mainly in the form of text and text representation as well as the IR system setting, e.g. database structures and retrieval techniques.

By applying the *cognitive viewpoint* as an epistemological foundation for IR research it becomes evident that one must take into account the variety of states of knowledge associated with these major participants in information retrieval interaction. Hence, IR interaction implies a cognitive holistic turn in IR research.

The book is organised with an initiating chapter describing the author's view of the emergence, scope and current state of information science as a scientific discipline. In order to correct the misunderstanding often observed in recent papers on the

philosophical basis for AI and IR, namely that cognitivism (or 'hard AI') is identical to the cognitive viewpoint advocated by Brookes, Belkin and the author during a decade, Chapter 2 provides an original epistemological analysis of the matter.

The discussion leads to a necessary re-conception of the understanding of the *concept of information* for information science. This understanding of information is a further development of N. Belkin's concept based on his ASK hypothesis (1977, 1978) as well as G. Wersig's earlier work on the issue (1971, 1973). It incorporates cognitive analyses of B.C. Brookes' Fundamental Equation for Information Science. The innovation of the concept lies in its explicit emphasis of conditions for both the sender *and* the recipient as to when we may talk of information associated with conveyed messages.

Based on a tri-partite *categorisation of IR research*, originally developed by the author (1988), Chapters 3–7 explore the R&D discussions and major results hitherto presented in the field. This framework is established according to the foci of research in IR: the system components and processes (the *traditional* approach, Chapter 4); the human participants and their information requirements (the *user-oriented* approach, Chapter 5); the integration of all the interaction processes taking place in IR (the *cognitive* approach, Chapters 6–7).

Chapter 6 discusses selected cognitive IR models and the user and intermediary knowledge characteristics and categories fundamental to the understanding of IR interaction. These models and typologies are based on empirical investigations and incorporate a fundamental distinction between various forms of IR knowledge and conceptual knowledge. In addition, the chapter provides an analysis of problems related to evaluation and relevance assessment, incorporating cognitive task modelling and information quality issues. Chapter 7 exhibits the characteristics of cognitive IR research, in particular by discussing its integrative properties and the role of knowledge-based IR. The approach is seen as an attempt to produce a synthesis concerning IR theory and application.

A central point for discussion is the degree to which an intermediary mechanism ought to carry out intensive user modelling, followed up by knowledge-based inference of relevant search strategies and retrieval of information, or alternatively, ought to make user modelling and inference instruments for supporting the user and the development of his information need and underlying problem, interest or goal. This supportive role of the intermediary in knowledge-based IR interaction implies the deliberate use of the user's own intelligence and associative capability, integrated with a high degree of transparency and structured feedback from IR systems as well as adaptive functionalities in the mechanism. In particular, the notion of *structured feedback* from (remote) IR systems plays a significant role. The feedback supports the user *and* the intermediary in their definition and understanding of the actual requirement for information, the underlying purpose as well as the entire performance in IR interaction. In short, the philosophy underlying this latter approach is to allow the participants to adapt to one another during IR interaction by means of supportive user modelling, integrated with modelling of, and adaption to, (remote) IR systems and information sources.

The author's position on this issue is the adaptive and supportive approach.

Drawing upon the results presented earlier in the book, Chapter 8 presents and discusses the consolidated framework of functional requirements for intermediary mechanisms in multi-domain and IR environments – the *MEDIATOR Model*.

This model is a further development and extension of the Monstrat Model's 10 functions and 23 tasks (Belkin, Seeger and Wersig, 1983; Daniels, Brooks and Belkin, 1985; Belkin et al., 1987) into 13 functional requirements and 54 sub-functions. It integrates the Monstrat Model's profound user-orientation with generalised domain and task knowledge as well as IR system adaption.

The Mediator Model displays three levels of functionality: a cognitive task modelling level; a cognitive adaption level; an IR effectiveness level.

At the first level, Mediator stresses the importance of long-term domain, user and IR system models, intended to be generated via field study analysis. At the second level, two functions are active and adaptive short-term model builders, geared toward the actual user and his goal and information need *as well as* the exploration of the potentialities in the IR environment. At the third level, the remaining eight functions are viewed as integrated structures concerned with request model building, the mapping of user profiles, the matching of retrieval strategies, and structured conceptual feedback generation, as well as their processes internal to the mechanism. In particular, its three-level distinction between pre-structured models, actual model building and functional performance, as well as the introduction of the Feedback, the IR System Adaptor, and the Domain Model functions, are seen as improvements of intermediary mechanism design in an adaptive and supportive knowledge-based IR environment.