The Turn

THE KLUWER INTERNATIONAL SERIES ON INFORMATION RETRIEVAL

Series Editor:
W. Bruce Croft
University of Massachusetts, Amherst

Also in the Series:

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation,*** *by Gerald Kowalski;* ISBN: 0-7923-9926-9

CROSS-LANGUAGE INFORMATION RETRIEVAL, *edited by Gregory Grefenstette;* ISBN: 0-7923-8122-X

TEXT RETRIEVAL AND FILTERING: ***Analytic Models of Performance***, *by Robert M. Losee;* ISBN: 0-7923-8177-7

INFORMATION RETRIEVAL: UNCERTAINTY AND LOGICS: ***Advanced Models for the Representation and Retrieval of Information***, *by Fabio Crestani, Mounia Lalmas, and Cornelis Joost van Rijsbergen;* ISBN: 0-7923-8302-8

DOCUMENT COMPUTING: ***Technologies for Managing Electronic Document Collections***, *by Ross Wilkinson, Timothy Arnold-Moore, Michael Fuller, Ron Sacks-Davis, James Thom, and Justin Zobel;* ISBN: 0-7923-8357-5

AUTOMATIC INDEXING AND ABSTRACTING OF DOCUMENT TEXTS, *by Marie-Francine Moens*; ISBN 0-7923-7793-1

ADVANCES IN INFORMATIONAL RETRIEVAL: ***Recent Research from the Center for Intelligent Information Retrieval***, *by W. Bruce Croft*; ISBN 0-7923-7812-1

INFORMATION RETRIEVAL SYSTEMS: ***Theory and Implementation,*** **Second Edition**, *by Gerald J. Kowalski and Mark T. Maybury;* ISBN: 0-7923-7924-1

PERSPECTIVES ON CONTENT-BASED MULTIMEDIA SYSTEMS, *by Jian Kang Wu; Mohan S. Kankanhalli;Joo-Hwee Lim;Dezhong Hong;* ISBN: 0-7923-7944-6

MINING THE WORLD WIDE WEB: ***An Information Search Approach***, *by George Chang, Marcus J. Healey, James A. M. McHugh, Jason T. L. Wang;* ISBN: 0-7923-7349-9

INTEGRATED REGION-BASED IMAGE RETRIEVAL, *by James Z. Wang;* ISBN: 0-7923-7350-2

TOPIC DETECTION AND TRACKING: Event-based Information Organization, *edited by James Allan;* ISBN: 0-7923-7664-1

LANGUAGE MODELING FOR INFORMATION RETRIEVAL, *edited by W. Bruce Croft; John Lafferty;* ISBN: 1-4020-1216-0

MACHINE LEARNING AND STATISTICAL MODELING APPROACHES TO IMAGE RETRIEVAL, *by Yixin Chen, Jia Li and James Z. Wang;* ISBN: 1-4020-8034-4

INFORMATION RETRIEVAL, ***Algorithms and Heuristics,*** *by David A. Grossman and Ophir Frieder,* 2nd ed.; ISBN: 1-4020-3003-7; PB: ISBN: 1-4020-3004-5

CHARTING A NEW COURSE, ***Natural Language Proceesing and Information Retrieval,*** *edited by J. Tait;* ISBN: 1-4020-3343-5

INTELLIGENT DOCUMENT RETRIEVAL, ***Exploiting Markup Structure,*** *by U. Kruschwitz;* ISBN: 1-4020-3767-8

The Turn

Integration of Information Seeking and Retrieval in Context

by

Peter Ingwersen
Royal School of Library and Information Science, Copenhagen, Denmark

and

Kalervo Järvelin
University of Tampere, Finland

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 1-4020-3850-X (HB)
ISBN-13 978-1-4020-3850-1 (HB)
ISBN-10 1-4020-3851-8 (e-book)
ISBN-13 978-1-4020-3851-8 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springeronline.com

Printed on acid-free paper

All Rights Reserved
© 2005 Springer
No part of this work may be reproduced, stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording
or otherwise, without written permission from the Publisher, with the exception
of any material supplied specifically for the purpose of being entered
and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands.

Acknowledgements

We wish to express our gratitude towards our colleagues and students for their encouragements as well as criticism, which we have been pleased to receive during the last five years. This was advantageous in developing our ideas, concepts, models and frameworks presented below.

We wish to acknowledge the support provided by the Nordic Research Academy (NORFA) 1999-2002, during which period Peter Ingwersen was visiting professor at Tampere University, and NORSLIS, the Nordic Research School in Library and Information Science, 2004-05. We also show our gratitude towards our Departments of Information Studies at the University of Tampere and the Royal School of Library and Information Science in Copenhagen for travel support as well as to our international friends, Marco Palmaci in particular, for backing us and making life more fun during the process.

In particular, we wish to thank Pia Borlund and Birger Larsen, Denmark, Jaana Kekäläinen and Pertti Vakkari from Finland, for their continuous and constructive discussions on central matters of concern regarding this book.

We are in particular in depths to our families for their continuing patience and encouragement.

Copenhagen and Tampere, May 2005

Peter Ingwersen & Kalervo Järvelin

Preface

There are many words we would have liked to write into the book title in order to describe the nature of this book. However, all the words did not fit in the title the reader already saw. We therefore chose *The Turn* as the main title to communicate the idea that it is time to look back and to look forward to develop a new integrated view of information seeking and retrieval: the field should turn off its separate narrow paths of research and construct a new avenue. Some of the elements for this avenue are given in the subtitle *Integration of Information Seeking and Retrieval in Context.* These words are intended to persuade a potential reader into the avenue project, preferably through the book. Some of many words that did not all fit into the subtitle are: the cognitive viewpoint; interaction between humans and machines; tasks, contexts and situations; knowledge acquisition, information seeking and use; models, frameworks and theories; research methodology. These are mentioned here because we see them as the foundation of the highway.

The present book takes off from the perspectives, ideas and findings presented in the monograph published thirteen years ago (Ingwersen, 1992). It reflects a further development of the cognitive viewpoint for information seeking and retrieval by providing a contextual holistic perspective. The quite individualistic perspective laid down in the former monograph is hence expanded into a social stance towards information behavior, including generation, searching and use of information. This stance is reflected in the proposed research framework and, in particular, associated with the further development of the conceptions of work and search tasks, information and information acquisition, IR interaction and the polyrepresentation principle. The monograph (1992, p. 123-201) saw IR interaction, the individual searcher's cognitive structures and the interface functionalities as important phenomena for study. The present book ascertains interaction as *the* central research issue in IS&R. The Mediator model with its thirteen basic interface functionalities (1992, p. 203) now represents but one of nine dimensions of our research framework. The monograph predicted a 'cognitive turn' for IR and a contextual IR theory development to take place in the time to come. The present book is/mirrors this turn and novel theoretical perspectives.

It is now the reader's privilege to assess whether or not the highway, as outlined through these words, looks promising enough for continuing beyond this point. Continuing means construction.

Peter Ingwersen and Kalervo Järvelin

Table of Contents

Acknowledgements ..... **V**

Preface ..... **VII**

Table of Contents ..... **IX**

1 Introduction **1**
1.1 Motivation and Intention 2
1.2 Laboratory-based Evaluation in IR 4
1.3 Some Problems of Laboratory-based Evaluation in IR 6
1.4 Extension of the Laboratory-based Evaluation in IR 9
1.5 Models in the Domain of IS&R 10
1.6 The Structure of the Book 16
1.7 Some Central Concepts 19
1.7.1 The Participants or Components of IS&R 19
1.7.2 The Tasks 19
1.7.3 The Contents 20
1.7.4 The Needs and Relevance 20
1.7.5 The Processes 21

2 The Cognitive Framework for Information **23**
2.1 The Cognitive View 23
2.1.1 The Central Dimensions of the Cognitive View 25
2.1.2 From an Individualistic Cognitive View… 28
2.1.3 … to the Holistic Cognitive View 29
2.1.4 The Role of Context in Individual Cognition 30
2.2 The Cognitive Information Concept in IS&R 31
2.2.1 Properties of Cognitive Models 34
2.2.2 Automatic Indexing: The Chinese Room Case 35
2.2.3 The Cognitive Free Fall – and Reconstruction 36
2.2.4 Other Central Information Conceptions 38
2.2.5 Relations to the Conditional Cognitive Information Concept 42

2.3 Implications of the Cognitive Information Conception 43
2.3.1 Information and Meaning 43
2.3.2 Basic Knowledge Types in IS&R 46
2.4 Information Acquisition: From Sensory Data to Scientific Discovery 47
2.4.1 Information Acquisition from Unknown Signs 48
2.4.2 Information Acquisition from Sensory Data: The Okawango Pursuit 51
2.4.3. Information Acquisition in Scientific Discovery 53

3 The Development of Information Seeking Research 55
3.1 Research Questions and Findings 56
3.1.1 Overviews of the Literature 56
3.1.2 Information Seeking Models 58
3.1.3 Conceptual Evolution 72
3.1.4 Empirical Findings 80
3.2 Research Methods 86
3.2.1 Research Strategies 87
3.2.2 Methods of Data Collection 89
3.2.3 Methods of Data Analysis 97
3.2.4 The Methodology in Non-Empirical Studies 101
3.3 Limitations and Open Problems 102
3.3.1. Theoretical Developments and Limitations 102
3.3.2. Empirical Developments and Limitations 107
3.3.3. Methodological Developments and Limitations 108

4 System-Oriented Information Retrieval 111
4.1 Models in System-Oriented IR Research 114
4.2 Major Developments of System-oriented IR Research 120
4.3 Documents, Requests and Relevance: Issues and Findings 123
4.3.1. Documents 123
4.3.2. Requests 126
4.3.3. Relevance Assessments 129
4.4 Indexing, Classification and Clustering: Issues and Findings 130
4.5 Interfaces and Visualization: Issues and Findings 135
4.6 Interaction and Query Modification: Issues and Findings 140
4.6.1. Relevance Feedback as Query Modification 141
4.6.2. Query Modification Based on Collection-Dependent Knowledge Structures 143
4.6.3. Query Modification Based on Ontologies 144
4.6.4. Structured Queries in Query Modification 146

4.6.5. Query Modification in Operational Systems Interaction Studies 150
4.7 Natural Language Processing: Issues and Findings 150
4.7.1. NLP Processing levels 152
4.7.2. Morphology in IR 153
4.7.3. Syntax and Disambiguation in IR 155
4.7.4. Semantics and Discourse in IR 158
4.7.5. Three Abstraction Levels of NLP in IR 160
4.8. Expert Systems and Interfaces for IR: Issues and Findings 162
4.8.1. IR Expert System Knowledge and Functions 163
4.8.2. I^3R – a Sample IR Expert System 166
4.8.3. IR Expert Systems Issues and Findings. 169
4.9 Research Methods 170
4.9.1. Laboratory Evaluation Studies 171
4.9.2. Operational Systems Evaluation Studies 174
4.9.3. The Interactive Track of TREC 177
4.10. Novel Test Collections and Performance Measures 178
4.10.1. Evaluation Measures: Relative Relevance and Ranked Half-Life 179
4.10.2. Evaluation Measures: Generalized Recall and Precision ... 180
4.10.3. Evaluation Measures: Cumulated Gain 182
4.11 Limitations and Open Problems 186

5 Cognitive and User-Oriented Information Retrieval 191
5.1 Models for Cognitive IR Research 195
5.1.1 Early Models of IR Interaction 195
5.1.2 Models for Cognitive IR Research 197
5.2 Cognitive IS&R Theory Building: ASK – Polyrepresentation: Issues and Findings 203
5.2.1 Taylor's Stages of Information Need Formation – and the Five Filters 204
5.2.2 The ASK Hypothesis 204
5.2.3 The Principle of Polyrepresentation 206
5.2.4 Empirical Evidence for Polyrepresentation 209
5.3. Searchers' Behavior, Cognitive Models and Styles – Issues and Findings 214
5.4. Standard Online IR Interaction – Issues and Findings 217
5.4.1 Search Tactics and Berry-Picking – Bates' Approach 218
5.4.2 Human Intermediary Behavior 219
5.4.3. End-user Online IR Interaction Studies 221
5.4.4 Explaining Online Search Effectiveness 222
5.4.5 Findings on Task-based IIR 224

5.5. Web IR Interaction 225
5.5.1 Large-scale Search Engine Studies 225
5.5.2 User-centred Surveys 227
5.5.3 Children Searching the Web 229
5.5.5 Digital Library Studies 230
5.6 Searcher-Associated Best Match IR Interaction 231
5.6.1 The Interactive OKAPI Experiments 231
5.6.2 The New Generation of Interactive Best Match Experiments 233
5.7 Relevance: Issues and Findings 234
5.7.1 The Early Relevance Research 234
5.7.2 The Dimensionality of Relevance 235
5.7.3 Analytic and Theoretical Relevance Developments 237
5.7.4 Central Empirical Relevance Studies 241
5.8 Research Methods 244
5.8.1 Variables and Strategies in Research 245
5.8.2 Data Collection Methods 247
5.8.3 Data Analysis Methods 248
5.9 Interactive IR Evaluation Methods: Simulated Work Task Situations 251
5.10 Central Achievements, Limitations and Open Problems 254
5.10.1 Limitations and Open Problems 256

6 The Integrated IS&R Research Framework 259
6.1 Building the Conceptual Framework 263
6.1.1 Authors of Information Objects 264
6.1.2 Adding Value to Information Objects – the Human Indexer 266
6.1.3 Cognitive Actors as Designers of Interfaces and IT Algorithms 269
6.1.4 Selectors – Actors Responsible for Availability and Access of Information Objects 271
6.1.5 The Information Seeker as Explorer of Sensory Data 273
6.2 Approaching the complexity of IS&R 274
6.2.1 The Information Seeker: The Central Actor in the Framework of IS&R 276
6.2.2 Situation at Hand and Models of Context 278
6.2.3 Work and Search Tasks 282
6.2.4 Knowledge Types in a Task Framework 284
6.2.5 Work and Search Task Complexity in IS&R 287
6.2.6 From Work Task to Information Need Formation in Search Tasks 289

6.2.7 The Label Effect 296
6.2.8 Interaction as the Vital Process in IS&R 300
6.2.9 Cognitive Use of IS&R Effectiveness Measures 304
6.3 Characteristics of the IS&R model: a summary 305
6.3.1 Major Characteristics of the Cognitive IS&R Framework 305
6.3.2 The Cognitive IS&R Framework as a Meta Theoretical Model 307
6.3.3 The Cognitive IS&R Research Framework: An Evaluation .. 309

7 Implications of the Cognitive Framework for IS&R 313
7.1 Design and Evaluation Frameworks for IS&R 316
7.1.1 IR Research in Isolation 318
7.1.2 Information Seeking Research in Isolation 320
7.1.3 The IS&R Design and Evaluation Framework 321
7.2 Variables of the Organizational Task Dimensions 324
7.2.1 Representation of Tasks for IS&R – Search Task Variables . 326
7.2.2 IR Research Questions: Capturing Task Types and their Representations 327
7.2.3 Information Seeking Research: Capturing Work Task Context 329
7.3 Actor and Perceived Task Dimensions of Variables 331
7.3.1 Polyrepresentation of the Actor's Cognitive Space 335
7.3.2 IR Research: Capturing Evidence on Searchers and Information Needs 335
7.3.3 The Assumption of Information Need Invariability 337
7.3.4 Information Seeking Research: Capturing Evidence from Searchers' Cognitive Space 339
7.4 Document and Source Types: Range of Variables 340
7.4.1 Polyrepresentation of Information Objects 342
7.4.2 IR Research: Capturing Evidence on Documents and Sources 343
7.4.3 Information Seeking Studies: Capturing Evidence of Documents and Sources 343
7.5 Algorithms and System Components: Dimensions of Variables 344
7.5.1 IR Research: Capturing Evidence on Algorithms and System Components 345
7.5.2 Polyrepresentation of Document Components and by Cognitive Space Statements 346
7.5.3 The Assumption of Feature (term) Independence 348
7.5.4 Information Seeking Studies: Capturing Evidence of Algorithms and System Components 350

7.6 Access Types: Variables of Interaction 351
7.6.1 IR Research: Capturing Evidence on Access and Interaction 352
7.6.2 The Assumption of Independent Relevance Assessments 353
7.6.3 Information Seeking Studies: Capturing Evidence on Access and Interaction 355
7.7 The Multidimensional Research Design Cube 356

8 Towards a Research Program 359
8.1 Information Objects, IT and Natural Task Contexts 361
8.2 Information Objects, IT and Actors 364
8.2.1 Alternative Research Question 367
8.3 Information Objects, Interface and Actors 368
8.4 Interfaces, Actors and Socio-organizational Contexts 370
8.5 Methodological Caveats and Summary 374

9 Conclusion 377

Definitions 381

References 393

Index 437