Amanda Spink
Charles Cole (Eds.)

New Directions in Cognitive Information Retrieval

Springer

New Directions in Cognitive Information Retrieval

THE INFORMATION RETRIEVAL SERIES

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 Processing and Information Retrieval, *edited by John I. Tait;* ISBN: 1-4020-3343-5

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

THE TURN: Integration of Information Seeking and Retrieval in Context, *by Peter Ingwersen, Kalervo Järvelin;* ISBN: 1-4020-3850-X

NEW DIRECTIONS IN COGNITIVE INFORMATION RETRIEVAL, *edited by Amanda Spink, Charles Cole;* ISBN: 1-4020-4013-X

New Directions in Cognitive Information Retrieval

Edited by

Amanda Spink

Faculty of Information Technology,
Queensland University of Technology,
Brisbane, Australia

and

Charles Cole

Graduate School of Library and Information Studies,
McGill University, Montreal, Quebec, Canada

Springer

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

ISBN-10 1-4020-4013-X (HB)
ISBN-13 978-1-4020-4013-9 (HB)
ISBN-10 1-4020-4014-8 (e-book)
ISBN-13 978-1-4020-4014-6 (e-book)

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

www.springer.com

Printed on acid-free paper

Printed in The Netherlands.

TABLE OF CONTENTS

V. Conclusions

PREFACE

Amanda Spink and Charles Cole

This book brings together chapters that present a range of new directions in theories, models and ideas about cognitive information retrieval (CIR). The aim of the book has been to draw out and examine new directions in CIR research. CIR is conceptualized as complex human information related human computer interaction processes that are embedded within an individual's everyday social and life context. CIR is an important part of the human information condition and critical to the development of new approaches to the design of Web and IR systems. Subsequently, CIR is an interdisciplinary area of study that includes research from information science, computer science, human factors, cognitive science, human computer interaction and other information related disciplines.

In this book we provide cutting edge interdisciplinary theories, models and directions in CIR research. The book does not provide a history of the field of CIR, but includes papers that represent potentially significant and new research into many aspects of human information behavior (HIB). Emerging frameworks, models and theories are providing a more complex view of CIR that includes multitasking, relevance feedback, HIB and visualization techniques, and longitudinal process models. In particular, the book includes papers by some outstanding and often less established researchers, and chapters that often challenge the established views and paradigms of CIR research. The final chapter of the book provides an initial integrated framework and an overview of the key trends, theories and models emerging in the field.

In this book, we focused on collecting papers that broaden and deepen the framework for our understanding of CIR. The editors invited authors to contribute chapters that represented emerging research directions and ideas, in an effort to build a framework that extends beyond existing models and research, and provide new directions for further research. The editors and chapter authors are drawn from the international boundaries of CIR research, and this international spread contributed greatly to the interdisciplinary nature of the chapters.

The book represents a major intellectual endeavor for the editors. The first editor (Spink) is an information scientist who has worked with, taught and has researched various dimensions of CIR since 1980. The second editor (Cole) is an information scientist who has worked with, taught and researched in various areas of CIR research since 1994. The editors have collaborated extensively over the last six years on researching CIR, including empirical and theoretical studies, journal and conference papers, and special journal issues.

The book is organized into five sections. Section I provides a brief introduction to the interdisciplinary field of CIR. Section II includes chapters that provide a discussion of

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* vii–viii.

key concepts for conceptualizing CIR, including relevance, cognitive, interactive and polyrepresentational approaches. Section III includes chapters that propose new approaches to the conceptualizing CIR processes within multitasking and HIB frameworks. Section IV includes chapters that discuss new CIR techniques, including relevance feedback, visualization and training frameworks. In Section V, the editors conclude the book by providing the range of new directions proposed in the chapters for further research.

Each section contains one or more chapters relating to the broader area of the section. Although each chapter is unique and comprehensive within its particular framework, and with its own reference list, the chapters are cross referenced where appropriate to illustrate how the different topics mesh together to form a broader expanse of CIR.

This book is intended as a resource for CIR researchers, educators, and practitioners. Researchers and students in the fields of information science, computer science, cognitive science, human factors and related disciplines, and scholars investigating CIR will all find the chapters presented here a valuable source of new ideas to help and further their respective research horizons. This book is also an appropriate text for undergraduate, graduate and doctoral level courses in areas of CIR and information science. Librarians, information professionals and others who provide information services will find this book useful for research-based insights into CIR. In addition, anyone who is interested in understanding the cognitive aspects of CIR will surely find this book a fascinating and enlightening read.

We greatly thank the chapter authors for their stimulating and ground breaking contributions. Many chapters represent the work of collaborations between researchers and their students. We also thank those who edited sections of the book, including Bhuva Narayan and Frances Alvarado-Albertorio from the University of Pittsburgh.

Amanda Spink, Faculty of Information Technology,
Queensland University of Technology
Charles Cole, Graduate School of Library and
Information Studies, McGill University

SECTION I

INTRODUCTION

CHAPTER 1

INTRODUCTION: NEW DIRECTIONS IN COGNITIVE INFORMATION RETRIEVAL

AMANDA SPINK
Queensland University of Technology, Australia

CHARLES COLE
McGill University

1. INTRODUCTION

Humans have used electronic information retrieval (IR) systems for more than 50 years as they evolved from experimental systems to full-scale Web search engines and digital libraries. The fields of library and information science (LIS), cognitive science, human factors and computer science have historically been the leading disciplines in conducting research that seeks to model human interaction with IR systems for all kinds of information related behaviors. As technology problems have been mastered, the theoretical and applied framework for studying human interaction with IR systems has evolved from systems-centered to more user-centered, or cognitive-centered approaches. However, cognitive information retrieval (CIR) research that focuses on user interaction with IR systems is still largely under-funded and is often not included at computing and systems design oriented conferences. But CIR-focused research continues, and there are signs that some IR systems designers in academia and the Web search business are realizing that user behavior research can provide valuable insights into systems design and evaluation.

The goal of our book is to provide an overview of new CIR research directions. This book does not provide a history of the research field of CIR. Instead, the book confronts new ways of looking at the human information condition with regard to our increasing need to interact with IR systems. The need has grown due to a number of factors, including the increased importance of information to more people in this information age. Also, IR was once considered document-oriented, but has now evolved to include multimedia, text, and other information objects. As a result, IR systems and their complexity have proliferated as users and user purposes for using them have also proliferated. Human interaction with IR systems can often be frustrating as people often lack an understanding of IR system functionality.

New more holistic directions in CIR are emerging that conceptualize human-IR system interaction at the human-computer interaction level (taking into account interface issues), the information behavior level (taking into account the role of IR system interaction in the total range of people's information behaviors), and the

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 3–10.

organizational/societal context (taking into account the information situation, task situation and problem situation of users when they use IR systems).

Researchers are thinking more broadly about information seekers rather than just as users of IR systems and the technical problems that may arise during the use.

In this book we also seek to highlight how the research field of CIR is more interdisciplinary and is now utilizing wider theories and models from sociology, psychology, communications, etc. The chapters in this book seek to stretch our understanding of CIR to incorporate questions of social and spatial factors, multitasking and non-linear dimensions, and relevance and learning issues.

The research in this book is above all conceptual, based on theory and model building. It is not descriptive (although descriptive studies of the user inform the conceptual research of many of the authors in this book).

Chapters in this book explore the conceptual operations, processes and information situation at the nexus between user and IR system, which informs the framework for conceptualizing the interaction between these two principal actors. The IR system is considered by the authors invited to contribute to this volume to be capable of assuming the role of an intermediary when the user utilizes an IR system to access information from the system's database. The role of intermediary was traditionally played by the librarian, who brought to the table knowledge of both cognition and the situation of their users, frequently on an intuitive or experiential level.

Alternatively, how do we build these human-librarian capabilities into IR system design to make these systems truly interactive? Although the perspective of this book is user cognition—the IR system must elicit from the user a special kind of input, creating interfaces that direct and organize the user's cognition so that the data accessed during the interaction has a true potential to be informational—our goal in developing this book has been to facilitate intellectual challenges to traditional IR system design assumptions and to champion new ideas.

These and other important research issues are explored in the following chapters.

2. BOOK OUTLINE

This book has five sections.

Section I: Chapter 1: Introduction

Chapter 1 provides an introduction to the book, setting out intentions and summaries of each chapter. We have divided up the chapters into sections, each with a theme.

Section II: CIR Concepts

Section 2 contains four chapters under the section title *Concepts*. These provide a definitional framework for the study of cognitive IR. Chapter 2 examines the concept of selection, which defines information use in cognitive IR interaction. Chapter 3 examines the concepts of polyrepresentation and the cognitive overlap of these representation states: the user's task, problem situation, cognitive, and information need states. Chapter 4 analyzes the concept of relevance, suggesting a differentiated perspective on the relevance concept be substituted for the unitary perspective currently used in IR design. And Chapter 5 examines the concepts of knowledge need and knowledge behavior as replacements for the concepts of information need and information behavior.

Chapter 2: Interactive Information Retrieval: Bringing the User to a Selection State—In Chapter 2, Cole, Beheshti, Leide and Large focus on the cognitive aspect of users as they interact with and react to environmental stimuli coming at them from the IR system. The authors closely examine other interactive IR models. Within a cognitive perspective, information need as a conceptual basis for the user in the interaction is problematic. Cole, Beheshti, Leide and Large then specify the terms and definitions of the user-system interaction they wish to examine in detail.

The authors seek to reconceptualize information need as the conceptual basis for user-system interaction. To this end, the interaction is represented by the authors as a series of interacting states: the user's cognitive state, the user's task or problem state, and the various need states for each task or problem that may arise during the user's interaction with the IR system. In turn, the IR system presents the user with information stimuli that have the potential to align the user's cognitive, task or information need states. The system stimulus message's role in the interaction is to stimulate the user to make a selection, thus putting the user in a Selection State. When the user is in a Selection State, an information channel forms leading to information that if found, understood and integrated into the user's knowledge structures, ends in an information process.

The last part of the chapter reports findings from a study of "selection" carried out in a naturalistic setting of history and psychology courses given at McGill University in 2003. The study tests the hypothesis that "selection" transforms the knowledge state of the subject. The study's findings are that there are indications that this may be the case.

Chapter 3: Cognitive Overlaps Along the Polyrepresentation Continuum—Chapter 3 by Birger Larsen and Peter Ingwersen describes a user and document representation methodology that gives the IR system a fuller picture of the user and the docment set than current simple request-based systems allow, based on the principle of polyrepresentation. For the user, the polyrepresentation consists of the user's various concurrent information needs, emotional states, tasks, organizational obligations etc.—a multipronged representation of the user's cognitive space. For the document set, the polyrepresentation consists of the citation links, thesaurus terms, selectors (e.g., journal name, etc.), indexers terms and author's headings, captions, etc.—a multipronged representation of the document set. The polyrepresentations are then represented in the algorithm used by the system to match the user with the IR system database's document set. The result is a cognitive overlap—of the user's cognitive space and the information space of various configurations of document representations from various search systems.

The hypothesis upon which cognitive overlap is premised is that representations of different cognitive and functional origins point to a set of information objects in a particular seeking context. The resulting cognitive overlap—a sort of elaborate Venn diagram with an intersection in the overlap—presents more relevant information objects to the user in the polyrepresentation of the user's cognitive space than each independent representation would be able to achieve. The potential of the cognitive overlap principle and polyrepresentation is therefore to serve as a common theoretical framework for research that integrates the information seeking or user perspective with the mainstream IR focus on designing and testing IR systems.

The authors outline research since the model was created that has tested various aspects of the model, particularly Larsen's own boomerang effect (forward citation) application of the model. The authors then indicate strengths and weaknesses of the model. One issue in implementing the polyrepresentation model is whether to utilize best or exact matching methods to achieve the cognitive overlaps. The authors present a polyrepresentation continuum to conceptually organize the main challenge for future directions for research based on the principle of polyrepresentation: to further identify flexible matching methods that can generate high quality cognitive overlaps from a variety of representations.

Chapter 4: Integrating Approaches to Relevance—Chapter 4 by Ian Ruthven proposes a multidimensional perspective on the concept of relevance. The relevance concept currently forms the basis of correction feedback to the IR system, allowing the system to revise its matching algorithm so that it more accurately matches the real—as opposed to the initially expressed—information need of the user, represented in the interaction by the user's query. We currently evaluate system performance according to the system's ability to retrieve topically relevant documents; we therefore build systems for retrieving topically relevant documents. This strong view of relevance, Ruthven believes—with a single, objective reality to relevance based on topic—has been a major obstacle to the development of more naturalistic methods and systems.

Unfortunately, information science, with its concern for the user, and the dominant, system-oriented school in IR research, continue as independent research traditions. System-oriented IR research ignores the extensive, user-based frameworks developed by information science and information science underestimates the influence of IR systems and interfaces within the information seeking process.

The author proposes that the relevance concept can be changed to reconcile the two research traditions. This new, differentiated view of relevance can be measured as relationships between different entities; for example, relevance can be measured against surrogates or whole objects, against a human cognitive need or against a concrete search request. Therefore, rather than the current, single, binary definition of relevance, differentiated relevance requires only that we are clear about what definitions of relevance we are utilizing in any investigation.

Chapter 5: New Cognitive Directions—Chapter 5 by Nigel Ford introduces new terms into cognitive IR research: Knowledge need and knowledge behavior. These new terms are necessary because the old terms, information need leading to information behavior, cannot be specified by the user, while knowledge need is the sought after knowledge, which can be specified by the user. Information need has long been known to be a problematic concept (Belkin, Oddy, and Brooks, 1982; cf. also Cole, Beheshti, Leide and Large, this volume), because the information stimuli and information processes that will generate the sought after knowledge cannot be predicted by the user. The realization and specification of the knowledge need, along with the evaluation of information behavior in terms of its contribution to satisfying the knowledge need, are termed here knowledge behavior.

The essential problem for IR systems then becomes the transfer of knowledge from the IR system to the user of these systems. Knowledge is only transferable to other contexts if there is a rich mapping of elements. Also, the processes occurring in information processes and knowledge transfer involve two sides of the brain operating in parallel. These two sides of the brain constitute two different components of human understanding, namely procedure-building and description-building. As a result, divergent information processing is needed in IR, not the present IR system which emphasizes convergent information processing.

The solution, Ford argues, is for the IR system to facilitate the user's creation of the threshold of knowledge about a topic needed to facilitate the required information processes leading to knowledge transfer; users need to become their own practitioner-researchers. Ford suggests the IR system enable the creation of dynamic, interactive, multi-perspective and flexible "evidence maps" that can facilitate the user's gradual build up of evidence into knowledge. The evidence maps would record the integration, cumulation, and disputation of—and other forms of interaction between—disparate information sources contained in the system's knowledge base and the user's gradually evolving knowledge of the topic through knowledge transfer from system to user.

Section III: CIR Processes

Section 3 contains three chapters under the section title *Processes*. These describe several processes involved in cognitive IR design and research practices. Chapter 6 concerns multitasking processes for cognitive IR design. Chapter 7 examines task process context variables that should be included in IR research. And Chapter 8 suggests that IR design should use as its guiding principle a social constructivism metaphor to encourage child learning when they use IR systems at home and at school.

Chapter 6: A Multitasking Framework for Cognitive Information Retrieval—Chapter 6 by Amanda Spink and Charles Cole analyzes user multitasking information behavior while the user is interacting with an IR system. Currently, IR systems require users to search sequentially and are largely designed to support limited types of searching based on specifying queries that select documents or Web sites to fulfill a single information task. However, IR system users naturally engage in multitasking while they are accessing information from an IR system. The users may begin their IR system interaction with multiple topics, or they may begin with a single topic and then develop additional topics during the search process. The authors define such behavior as natural, and examine possible mechanisms that enable the user to switch from one task to another while ostensibly engaged in accessing information from an IR system for one topic or task.

The authors begin the process of modeling multitasking within a cognitive IR framework, starting from Saracevic's (1997) Stratified model of user-IR system interaction. They define the central problem as the user coordinating between various levels of problem identification, system problems, etc. for a single topic or task—on the level of searching the IR system—and the user's wider thinking while interacting with the IR system—on the level of seeking information, which involves coordinating multiple search tasks and multiple topic tasks.

Chapter 7: Explanation in Information Seeking and Retrieval—In Chapter 7, Pertti Vakkari and Kalervo Järvelin examine task as a unifying variable in IR research. Presently, IR research originates from three separate fields: information science, computer science and contributions from the field of social science. However, the three communities do not really communicate with each other, and use different methodologies. The authors examine these different methodologies in terms of dependent, independent and controlled for variables. The results of their examination suggest that the research communities within these three areas focus on different phenomena in terms of the independent and, in particular, the dependent variables used in the studies.

Because IR serves the goals of seeking, and information seeking primarily serves the goals of the user's work task (or other interest), the authors suggest a possible area of reconciliation is the context or task for the user's information search. The authors propose ways of increasing the overlap between the three competing research fields by enriching basic research frameworks with research variables related to the user's work tasks as well as the user's cultural activities/interests.

Vakkari and Järvelin examine two studies that can be seen to utilize a task-based approach to information seeking and retrieval research: a study investigating the use of electronic high school learning materials when answering high school exam questions, and an investigation that compares, via (among others) outcome variables, experts and novices engaged in business problem solving.

Chapter 8: Towards An Alternative Information Retrieval System for Children—Chapter 8 by Beheshti, Bowler, Large and Nesset examines the complete picture of IR research for children, framing the discussion in four research traditions based on four different metaphors: behaviorism, information processing, cognitive constructivism and social constructivism. The information processing metaphor is still the dominant metaphor in IR design. According to this IR metaphor, information seekers could, it was assumed, negotiate through the universe of information with ease, once their own mental models matched those of the experts. External forces (external to the specifics of information searching), such as the user's motivation, anxiety, or influences from the larger society were ignored.

In contrast to IR research and its information processing metaphor, in learning and information seeking (IS) research the current metaphor is social constructivism. Because of these different metaphor frameworks, the central problem the authors address in this chapter is reconciling the divide between IR and IS research regarding information seeking and searching of IR systems. Especially for children, information seeking and interaction with an IR system is done within a specific social context (social constructivism), while information processing at the individual level is about cognition (cognitive constructivism). The authors ask: How can the concept of a community of discourse be transferred into information technology where individuals seemingly work in solitude while they search for information? The question suggests a radical overturn of the way we think about IR, particularly for children.

The authors suggest a reconciliation research framework based on a social constructivism metaphor: the individual constructs learning in a particular social context. The affective factor, where the system gives children emotional and empathetic-tutoring dimensions, should be built into the interaction. The authors propose an alternative IR system for children which applies the concept of a Zone of Proximal Development—a social constructivist concept—to information seeking, an activity that is essentially conducted alone using an electronic data base. This alternative IR system involves implementing a supportive emotional context better adapted for children using learning-based IR systems.

Section IV: CIR Techniques

Section 4 contains three chapters under the section title *Techniques*. These describe the techniques involved in cognitive IR. Chapter 9 examines implicit relevance feedback techniques—the metrics used and assumptions researchers make about user intention from observing user behavior when they interact with IR systems. Chapter 10 describes visualization techniques for representing knowledge domains that facilitate knowledge integration for domain novice users of IR systems. Chapter 11 describes techniques of user search training that can be built into cognitive IR system design.

Chapter 9: Implicit Feedback: Using Behavior to Infer Relevance—Chapter 9 by Diane Kelly examines the techniques used by researchers to study the implicit relevance feedback behavior of users engaged in interacting with an IR system. The problem confronted by Kelly is users not engaging in the relevance feedback activities currently provided by IR systems. Implicit relevance feedback techniques may offer some solution to this problem. Implicit relevance feedback techniques unobtrusively obtain information about users, their needs and document preferences, by watching their natural interactions with systems.

The author refers to her own previous work which observed user behavior then classified the behavior into five categories describing the underlying intent of the observed behavior. The resulting five categories are: examine, retain, reference, annotate and create. Kelly then reviews the implicit relevance feedback research to illustrate how such studies have been conducted and how feedback has typically been measured and used. The author concludes that current studies of implicit feedback devote little time to exploring the validity and reliability of the metrics they employ, assuming that the metrics represent the actual behavior of interest (e.g. the length of time a document is displayed in the browser window is assumed by these studies to be equivalent to how long the user reads the document).

The chapter ends with an identification of the challenges for future implicit relevance feedback research.

Chapter 10. Educational Knowledge Domain Visualizations: Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise—Chapter 10 by Peter Hook and Katy Börner analyzes various visual representation techniques for the purpose of improving the useability of Educational Knowledge Domain Visualizations (KDVs). KDVs visualize information spaces for specific topics or a broad range of topics, indicating to the user some sort of structure in a given information space. They can be incorporated into the IR system either at the front-end, as an overview of a topic space to assist the user in formulating an access point to the system's database, or at the backend of the system in the results list. The visual representation of a topic or domain structure assists the user, especially the domain novice, in discerning the most relevant entry points into the information space, as well as facilitating user integration of the objective knowledge space represented by the KDV into the user's subjective image of it. KDVs facilitate the user's search task of organizing the information search, information access, and management of his or her interaction with the IR system, and the information found in the system's database.

The authors are primarily interested in how visualizing knowledge domains can facilitate user internalization of scholarly knowledge by taking advantage of the way human vision and human spatial cognition work. From this point of view, the chapter outlines best-practice visualization design, then discusses and compares cartographic maps, including thematic maps, cognitive and concept maps, Knowledge Domain Visualizations, as well as Henry Beck's 1933 famous metro map. Henry Beck's 1933 redesign of the London Underground (subway) map was a revolutionary advancement in graphic design that has recently begun to serve as a model for the effective communication of system information spaces to domain novices and domain experts alike.

Chapter 11. Learning and Training to Search—In Chapter 11, Wendy Lucas and Heikki Topi describe techniques for training users in the information searching of IR systems. The authors set these search training techniques inside the broader perspective of information seeking. Different stages in the seeking process lead to different user cognitive processes, requiring different training models. Within each seeking stage, the user goes through a search process. Stage 1 in the Lucas and Topi model of the search process is the articulation of the user's information need. Stage 2 is the conceptualization of the need as an executable query. Stage 3 is the formulation of the conceptual query for a given IR system. Stage 4 is entering the query into the search interface. And Stage 5 is the user understanding and interpreting the search results. Each stage requires different information search training procedures, which are, in turn, different depending on the stage of the seeking process the user is in at the time of the interaction with the IR system.

The second Lucas and Topi search training strategy is to gather user errors and then specifically train users on correcting these errors. Categories of user errors are conceptual knowledge errors, semantic knowledge errors, and technical knowledge errors (Sormunen and Pennanen, 2004). Examples of user errors from the Excite study of Spink, Wolfram, Jansen, and Saracevic (2001) are the infrequent use of Boolean operators and errors when the Boolean operators are used, such as errors in capitalization, the use of the ampersand instead of AND, etc.

Section V: Chapter 12: Conclusion and Further Research

Chapter 12 provides a conclusion and a discussion of further research.

REFERENCES

Belkin, N. J., Oddy, R., and Brooks, H. (1982). ASK for Information Retrieval. *Journal of Documentation*, 38(2): 61–71.

Saracevic, T. (1997). The Stratified Model of Information Retrieval Interaction: Extension and Applications. In *Proceedings of the Annual Meeting of the American Society for Information Science,* 34, 313–327.

Sormunen, E. and Pennanen, S. (2004). The Challenge of Automated Tutoring in Web-based Learning Environments for Information Retrieval Instruction. *Information Research-An International Electronic Journal,* 9(2), Art. No. 169.

Spink, A. Wolfram, D., Jansen, B. J., and Saracevic, T. (2001). Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science,* 53(2), 226–234.

SECTION II

CIR CONCEPTS

CHAPTER 2

INTERACTIVE INFORMATION RETRIEVAL: BRINGING THE USER TO A SELECTION STATE

CHARLES COLE, JAMSHID BEHESHTI, JOHN E. LEIDE, AND ANDREW LARGE

Graduate School of Library and Information Studies, McGill University

1. INTRODUCTION

There have been various approaches to conceptualizing interactive information retrieval (IR), which can be generally divided into system and user approaches (Hearst, 1999; cf. also Spink, 1997). Both system and user approaches define user-system interaction in terms of the system and the user reacting to the actions or behaviors of the other: the system reacts to the user's input; the user to the output of the system (Spink, 1997). In system approach models of the interaction, e.g., Moran (1981), "[T]he user initiates an action or operation and the system responds in some way which in turn leads the user to initiate another action and so on" (Beaulieu, 2000, p. 433). In its purest form, the system approach models the user as a reactive part of the interaction, with the system taking the lead (Bates, 1990). User approaches, on the other hand, in their purest form wish to insert a model of the user in all its socio-cognitive dimensions, to the extent that system designers consider such approaches impractical (Vakkari and Jarvelin, 2005, Chap. 7, this volume). The cognitive approach to IR interaction attempts to overcome this divide (Ruthven, 2005, Chap. 4, this volume; Vakkari and Jarvelin, 2005 Chap. 7, this volume) by representing the cognitive elements of both system designers and the user in the interaction model (Larsen and Ingwersen, 2005 Chap. 3, this volume).

There are cognitive approach researchers meeting in a central ground from both the system and user side. On the system side, are computer scientists employing cognitive research to design more effective IR systems from the point of view of the user's task (Nathan, 1990; Fischer, Henninger, and Redmiles, 1991; O'Day and Jeffries, 1993; Russell et al., 1993; Kitajima and Polson, 1996; Terwilliger and Polson, 1997). On the user side are cognitive approach researchers applying methods, concepts and models from psychology to design systems that are more in tune with how users acquire information (e.g., Belkin, 1980; Ford (2005, Chap. 5, this volume); Ingwersen (Larsen and Ingwersen, 2005, Chap. 3, this volume); Saracevic, 1996; Vakkari (Vakkari and Jarvelin, 2005, Chap. 7, this volume)).

One cognitive meeting ground between user and system orientations is the conceptualization of the user-system interaction in terms of agents representing the user and the system acting on behalf of each actor in the transaction. The agents also have some sort of effect the one on the other that is not negligible. For example, Human-Computer Interaction (HCI) researcher Storrs (1994, p. 173) defines interaction as the

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 13–41.

"exchange of information between participating agents through sets of information channels (interfaces) for the purpose of altering their states."

The agent for the IR system is evolving but is centered on the results list, summaries of the results list, visualizations of the results list (Hook and Borner, 2005, this volume; Markoff, 2005), or any other representation of documents, Web sites, etc. contained in the system's database and deemed by the system to match the user's information need as it is represented to the system by the user's agent in the transaction, the query. The user's query represents the user's information need in system approach IR research. However, this view of information need as the conceptual basis for the user's agent in the user-system interaction has been questioned by cognitive approach research.

In this chapter, after describing the problem the cognitive approach has with information need, we propose a model of the user-system interaction that eliminates the user's information need as the basis for the user's initial request statement to the system. Instead, we reconceptualize the interaction using terms that are common to both user and system approaches to interactive IR design, such as *state* and *selection*. The term *state*, to represent some state of the user's mind or thinking, has often been used in cognitive approach IR system research. Selection or selecting is an even older term in considering the user in the design of IR systems. In the pre-Internet, library environment, the trend towards open stacks and self-service in the late 19th century (Cutter, 1891–1893, p. 5) caused catalogers, indexers and classificationists to devise systems that assisted the user to (i) find, (ii) identify and (iii) select needed information while using the library's IR system (the card catalog). The three objects as they were called (i.e., objectives) were expressly codified in 1876 by Cutter (1876/1904) in the introduction to his Cataloging Rules for a Dictionary Catalog.

> (A 4th object, the obtaining object, was added in 1997 by the International Federation of Library Associations and Institutions (IFLA); and for the interactive technology era, Svenonius, 2000, has recently added a 5th, the navigation object. We discuss the 5th "navigation" object in Leide et al., 2003. The 2nd object, the collocation or identification object is discussed in Cole et al., 2005.)

The 3rd object, the selection object, is the subject of this chapter. In the next section, we discuss the weakness of information need as the conceptual basis for the user's agent in IR interaction.

1.1. Information Need as a Basis for User's Agent in Interaction

While both user and system approaches have the same goal—to maximize the utility to the user of the information found by the system—user oriented researchers have long believed (since for example, Belkin, Oddy, and Brooks, 1982) that the problem with current IR systems is their unidimensional model of the user as he or she is represented in the interaction between user and IR system, particularly the assumption that the user's request statement to the system encapsulates the user's information need.

In system approaches, the user's agent in the interaction, the query, is assumed to be an apt representation of the user's information need. Based on this assumption, modification of the need statement as a result of system feedback to the user (via interim results, thesauri, or other search aids offered by the system to help the user reformulate the query) is to correct communication errors between user and system—i.e., to refine the match between the user's information need and the system output. "Such reformulation strategies assume that the user's information need is fixed throughout the search session, and a dialogue of sorts is taking place between the system and the user to help the system match the user's fixed need" (O'Day and Jeffries, 1993, p. 440). Hearst (1999) and Ingwersen (1992) diagram this simplistic view of user-IR system interaction based on a unidimensional view of the user's information need.

In this unidimensional, static view of information need, the user has a topic and knows what information he or she needs and types in topic elements as the request statement to the system. We currently evaluate system performance according to the system's ability to retrieve topically relevant documents; we therefore build systems for retrieving topically relevant documents (Ruthven, 2005, Chap. 4, this volume).

Belkin and colleagues (Belkin et al., 1995) have long believed (Belkin, Oddy, and Brooks, 1982) that asking users to specify their information need in a query to the system is unrealistic. That in the real world, "Studies have shown that a searcher's ability to get desired results is often very weak (Spink et al., 2001; Lucas and Topi, 2002). Key among these findings is the inability of searchers to choose correct search terms . . . " (Lucas and Topi, 2005 Chap. 11, this volume).

Rather than attempting to represent their information need in a query, user oriented researchers believe a more realistic approach is to bring the user's problem or problematic situation into the interaction. The user's problem situation has been an important Library and Information Science (LIS) concept beginning with Wersig (1971). It is primarily an information seeking concept as it includes derived user information behavior that is independent of the IR system—i.e., it is not purely an information search concept concerned only with user-system interface issues (Wilson, 1999; 2000).

By bringing the problem situation of the user into the user-system interaction, there is an immediate addition of a sociological perspective to the user's agent in the transaction between user and system, particularly a consideration for the user's task when seeking the information. The user's problem situation also includes consideration of the organizational constraints on the user, and the importance of using the information found during the transaction to communicate a message via the user's task to the organizational milieu governing successful performance of the task. The challenge in bringing these wider user concerns into the user-system interaction is how to ascertain then represent the problem situation in the user's agent, a challenge taken up by user approaches to IR interaction since Belkin, Oddy, and Brook's (1982) ASK concept (for an analysis of the ASK concept, cf. Cole et al. in press).

The general problem addressed in this chapter is to devise a user or cognitive approach to modeling the user's problem situation so that it can form a substitute for information need as the user's agent in user-system interaction. In Section 2, we review the literature on cognitive approach multidimensional models of user-system interaction, which consider the user's problem situation in the interaction. In Section 3, we define terms for the proposed model, then diagram the model. In Section 4, we state the specific problem of the study which begins to test certain aspects of the model. In Section 5, we report the study. We end the chapter in Section 6 with conclusions and future directions for our research.

2. LITERATURE REVIEW: COGNITIVE IR'S MODELING OF THE USER'S AGENT

A number of multidimensional approaches to creating a viable agent for the user in the user-system interaction have been proposed by user oriented researchers (cf. Jansen, Spink, and Pedersen, 2005). An important distinction in these models is the difference between information search and seeking. The latter models set out to bring more general information behaviors of the user into the transaction, information behaviors that are conditioned by the user's task, the purpose of the task, organisational constraints, etc. when he or she uses an IR system to seek information. Considerations of the user's task and the wider sociological implications of the user's search center on the decades old (since Wersig, 1971) information seeking term "problematic situation" or "problem situation."

Beaulieu (2000) distinguishes information search interaction models from information seeking interaction models. Examples of information search-based models are Bates (1990), who focuses on the tactics and strategies of the user's search of the IR system; and Lalmas and Ruthven (1999), who distil interaction to the system teaching the user how to conduct the search (cf. also, Lucas and Topi, 2005, Chap. 11, this volume). Examples of information seeking-based models are Efthimiadis and Robertson (1989), who divide user-system interaction into stages of the information seeking process; Vakkari and Jarvelin (2005, Chap. 7, this volume), who focus on the user's situation or problem task in seeking information via an IR system; and Larsen and Ingwersen (2005, Chap. 3, this volume), who propose a polyrepresentation IR interaction model which includes a wider notion of the user's problematic situation.

A certain number of multidimensional approaches to interaction propose operationalizing dimensions in terms of levels. Bates (1990), for example, divides interaction into four search activity levels: moves, tactics, stratagems and strategies. Spink and Cole (2005, Chap. 6, this volume) utilize Saracevic's (1996) Stratified Model of user-system interaction in their of description of multitasking in cognitive IR (cf. also, Spink and Saracevic, 1997).

Saracevic's Stratified Model of user-IR interaction posits a multidimensional view of user-IR system interaction, expanding the two actor traditional model of the user on one side and the computer IR system on the other, by dividing the user and computer into strata or levels. All the levels interact with each other and are affected or influenced by the interaction taking place between the other strata. The computer side is made up of four strata: the surface, engineering, processing and content levels; while the user side is made up of the surface, cognitive, situational and affective levels. The surface level interaction corresponds to the traditional model of user-IR system interaction described above, but also takes into account user strategies, tactics and other processes where the user explores the surface features of the IR system, or deals with system visualizations of the results list, for example, again on the surface or interface level of the computer side. The cognitive level, however, explores the user interacting with, for example, the content level of the computer side, assimilating texts cognitively. The situational level on the user side takes into account the task or problem situation of the user, which describes, along with the affective level on the user side, the panoply of motivations driving the user to seek information via the IR system.

Spink and Cole (2005, Chap. 6, this volume) in their model of multitasking during user-system interaction, conceptualize interaction between Saracevic's Stratified Model levels as a human information coordinating behavior (HICB), a mechanism for task switching, for both search and information seeking tasks. An approach that emphasizes the affective level of the interaction is Beheshti et al., (2005, Chap. 8, this volume), who conceptualize a social constructivist IR system that adds emotional support to the user-system interaction for the design of IR systems for children users. Saracevic's Stratified Model provides a much more complex view of user-IR system interaction than the traditional model of user-IR system interaction outlined in the Introduction above. Contrary to Beaulieu (2000), we place Saracevic's Stratified Model of interaction in the seeking category because the user's situational level is so strongly considered in the model.

A model that explicitly brings information seeking into the user-system interaction is Belkin et al.'s (1995) Information Seeking Strategies (ISS) Model of Interaction. The ISS Model is divided into four dimensions:

- Method of Interaction (search or scan)
- Goal of Interaction (learning or selecting)
- Mode of Retrieval (specification or recognition)
- Resource Considered (information or meta-information)

Any user's ISS can be categorized depending on where it is placed on all four of the dimensions, giving (if dichotomous values) a total of 16 different ISSs. The strategies "arise from characteristics of the person's problematic situation (Wersig, 1979); in particular, the user's state of knowledge and information-seeking goals" (Belkin et al., 1995, p. 380). Within one search session, the user exhibits different ISSs "as [the] person's problematic situation changes" (Belkin et al., 1995, p. 381).

The interaction between system and user in the ISS Model is conceived of as a dialog, with dialog states and action states, shown in a diagram of the "discourse act level" as circles and squares respectively. If for instance the user makes a request to the system for information, the user is in a request state (a circle). The system provides the information, it is accepted or rejected indicating a terminal state, or begins negotiation with the system indicating a dialog state.

The ISS Model terms "state" and "selection" are used by both user and system-oriented researchers in IR interaction. For system approach utilization of the word "selection" cf. Shannon, 1949; for system approach utilization of the word "state," for example in the Turing machine, cf. Herken, 1995. Unusually for Belkin, the ISS Model slants usage of these terms in the direction of the system approach. The user approaches employ a broader definition, anchored in a multidimensional perspective of human information behavior (cf. Larsen and Ingwersen, 2005, Chap. 3, this volume). The next sections discuss user-approach definitions for these and other terms.

3. TERM DEFINITIONS

3.1. State

The user or cognitive approach utilization of the word "state" is a developing concept in cognitive information retrieval. Belkin, Oddy, and Brooks (1982) utilize the word "state" in their central ASK concept (Anomalous States of Knowledge), to describe the states that lead users to seek information, which includes searching an IR system. Ingwersen (1996) uses the term cognitive state of the user, which includes interaction with the user's state of uncertainty. In this theory, the user's cognitive state is transformed through interaction with potential information objects from an IR system. Wilson (1999) also uses the term cognitive state (in addition to the user's physiological and affective states); in the Wilson model, the cognitive, physiological and affective states change depending on the user's stage of information seeking. Saracevic (1996) defines IR interaction's purpose as to "affect" the user's cognitive state: "The IR interaction is then a dialogue between the participants—user and 'computer'—through an interface with the main purpose to affect the cognitive state of the user for effective use in connection with an application at hand" (Saracevic, 1996, p. 7).

The term "state" describes a static position of a person or thing (Oxford Dictionary, 1984), thus its utility for both use and system approaches. The state for the system is "predefined and fundamentally static" (Ingwersen, 1996, p. 7). However, to describe the user side of the interaction, the preferred term in user approaches because of its dynamic quality is "cognitive state," which is defined as: "the state of the person's cognitive processes" (WordWeb Online, http://www.wordwebonline.com).

The purpose of utilizing the term "cognitive state" is

- to replace the problematic term "information need,"
- to introduce a process conceptualization to the user's agent, thus replacing information need which is static in the traditional model of user-system interaction, and
- to expand the conceptual sophistication of the user's agent in the user-system interaction—i.e., the user's query to the system—to include considerations of the user's problematic situation.

As a result of this substitution of terms: "in addition to the traditional request formulation, supplementary separate paths for exploration and retrieval are made available: paths which otherwise might never be reached via the need-associated query versions alone" (Ingwersen, 1996, p. 39).

There is thus more than one cognitive process in the term "cognitive state." Ingwersen (1996, p. 41), however, who also uses the term cognitive state, prefers using the term cognitive *space* to model the multitude of cognitive processes involved in the user-system interaction: "The concept of polyrepresentation of the user's cognitive space involves representing not only the current (often topic) information need, but also (and more importantly) the underlying problem space, actual work task or interest, and the dominant work domain(s)" (Ingwersen, 1996, p. 41). As Ingwersen is a contributor to this volume (Larsen and Ingwersen, 2005, Chap. 3, this volume), we will briefly describe what is meant by "polyrepresentation of the user's cognitive space."

Chapter 3 of this volume by Birger Larsen and Peter Ingwersen describes a user representation methodology that gives the IR system a fuller picture of the user than current simple request-based systems allow, based on the principle of polyrepresentation. The polyrepresentation consists of the user's various concurrent information needs, emotional states, tasks, organizational obligations etc.—a multipronged representation of the user's cognitive space. During interaction between user and system, the system captures these representations via a request model builder (RMB) interface. The cognitive overlap of the polyrepresentation is then represented in the algorithm used by the system to match the user's cognitive space with the IR system database's document set, creating an information space for the particular user session.

The hypothesis upon which cognitive overlap is based is that representations of different cognitive and functional origins point to a set of information objects in a particular seeking context which is different than the information objects now pointed to in current systems by a user request statement alone, one based only on information need.

The RMB is derived from only three of the four cognitive structures inside the user's cognitive space:

- the user's information need statement,
- the user's problem and goal statements,
- the user's work-task/interest descriptions.

The three representations in the polyrepresentation representing the user's cognitive space are "ANDed" together in the system's matching algorithm, creating something like a Venn diagram. However, we are particularly interested in the four cognitive structures that feed into the three representations considered by the RMB in its formulation of the polyrepresentation of the user's cognitive space:

- Information Need
- Problem Space
- Current Cognitive State
- Work-Task/Interest Domain

Note that the only structure in the above list which does not have a separate representation in the cognitive space polyrepresentation is the user's Current Cognitive State.

We wish to further define then model the user's Current Cognitive State. The Current Cognitive State (CCS) is defined by Larsen and Ingwersen (Chapter 3, this volume) as the "(limited) awareness or sense of what is desired." This awareness is variable, changing throughout the user's search session. The CCS is distinct from the user's information need, which is the stated need. As all four cognitive structures are interconnected, the CCS is modeled in conjunction with the other three cognitive structures.

In our model (see below), we start with Larsen and Ingwersen's statement that the three variable structures that constitute the user's cognitive space "are easily affected by <u>external input and/or thinking processes,</u> while the work task/interest perception is set in a social context and may be more stable" (emphasis added). Also, "The problem space corresponds to Belkin's ASK (Belkin, 1980; 1982) and is separated from the information need because the same problematic situation may give rise to several different information needs" (Larsen and Ingwersen, 2005, Chap. 3, this volume).

3.2. Current Cognitive State (CCS)

We begin modeling Current Cognitive State (CCS) by focusing on the Larsen and Ingwersen distinction between the user's stated information need and the desired information. The latter only is included in their term CCS, but the other structures in the user's cognitive space, including information need, are interconnected. Looked at this way, the distinction between the stated and desired information need invokes Taylor's (1968) division of information need into four levels.

Taylor's (1968) classic article on information need describes the user's actual information need as being unconscious (Level 1). A modified form of the need is vaguely conscious but nevertheless unspecifiable (Level 2), but even the specifiable form of the need (Level 3) is not the one used by the person when accessing information (i.e., asking a librarian or IR system a question). The user accesses information from a librarian or system employing compromised language or word tokens (Level 4) as the query to the IR system because he/she thinks the information need will not be processed by the IR system unless it has been "appropriately" coded.

From this model of information need, the stated information need of the user is "a statement of what the user does not know" (Belkin, Oddy, and Brooks, 1982, p. 64). Belkin, Oddy, and Brooks (1982) define their ASK as Taylor's level 1 or 2, not, as Larsen and Ingwersen (2005, Chap. 3, this volume) have done, which is to associate the ASK concept with another structure entirely in the user's cognitive space, the user's problem space.

Let us operationalize Larsen and Ingwersen's CCS and Taylor's Level 1 as the user's goal state in seeking information using an interactive IR system, which is entirely different from the user's request statement—i.e., the user starts to seek information using an IR system by typing in a request statement that, in order to be successful, must find information in the IR system's database he or she does not yet know. As we are conceptualizing the information seeking in an interactive dialogue between user and system, the user's information need request (Taylor's Level 4) is a start state while Level 1 is the goal state of the user-system interaction. The system, therefore, must interact with the user, via the user's agent the information need request statement, to bring the user from a start state to the goal state.

This conceptualization of the role of user-system interaction agrees with the long held role of the librarian as a transducer, someone who must think in terms of the answer to the user's question in order to find appropriate sources of information for the user (Pomerantz, 2005). Taylor's (1968) five filters in fact focus mainly (4 out of the 5) on the librarian's calculation of the user's answer, not the question or request statement (Cole, Kennedy, and Carter, 1996). The librarian must do this in order to locate information sources where the answer to the user's question can most likely be found (Katz, 1997; Bopp and Smith, 2001). (For a discussion of classification schemes of user answers based on user questions, cf. Pomerantz, 2005.)

We have so far operationalized the user's question as the request statement or query to the system, which for now we will call the user's start state. The goal state is finding or locating representations of information objects contained in the system's database that will answer the user's question, which we call here the user's goal state. The purpose of user-system interaction is "the process(es) which are mental, physical, or perceptual activities required to move from the initial [start] state to the goal state" (MacMullin and Taylor, 1984, p. 95).

Figure 1 diagrams the start and goal states of the user-IR system interaction, inside a large CCS box. The CCS box represents the current cognitive state of the user as he or she is interacting with the IR retrieval system. The start state has something to do with the user's request or query statement, which is a question. The goal state is

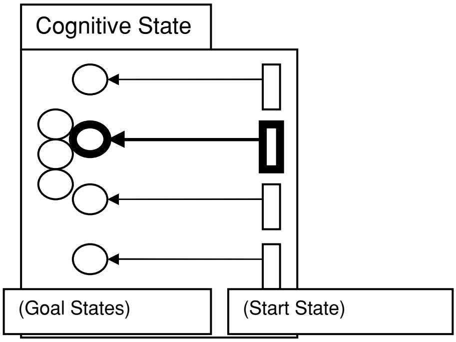

Figure 1: Cognitive state diagrammed as start and goal states

the user's question to the system that will procure system output that will satisfy the user's information need—i.e., the system output (results list) will facilitate the user formulating an answer to the question contained in the goal state query.

The purpose of the user-system interaction is to alter the user's state, to get the user from a start state to a goal state, so that the user is in a state to find information that will satisfy the user's information need. As the information need cannot be expressed in the request statement at the beginning of the interaction, because the user does not know what his or her information need is at the start of the session, this puts the system in an awkward position: it must somehow, through interacting with the user's agent, the user start state query, find and present to the user information from its database that will:

1. get the user to the goal state; and
2. satisfy the user's information need tied to the user's goal state.

3.3. Goal State

We have referred earlier to the user or cognitive approach to IR interaction which emphasizes the information seeking context rather than the information searching context of the traditional system approach to IR system design. An information seeking conceptualization of information need embeds information need and seeking in the activity that generates it—i.e., it considers the user's purpose or task in seeking the information. As previously discussed, the concept of information need in information seeking research immediately takes on a wider, dynamic level of the user's problematic situation.

Vakkari and Jarvelin (2005, Chap. 7, this volume) have operationalized the user's purpose in seeking information as the user's task inside an organizational work task perspective. The costs of the IR system, therefore, "should be justified by the quality of the deliverables for some necessary and important work tasks" (Vakkari and Jarvelin, 2005, Chap. 7, this volume).

In this view, the actions supported by the IR system are those that enable the user to find information for the performance of the user's task rather than the immediate actions concerned with searching an IR system (i.e., making the IR system work).

The task-based approach to information seeking, information need, and interactive IR design is still a minority concern, however (for an analysis of task-based information seeking research, cf. Vakkari, 1999).

In the task-based view of the user and the information seeking conceptualization of the user's information need, the goal state of the user is readiness or ability to interact with the IR system for the purpose of finding information that will be useful to the performance of the task for which information is being sought. The purpose of the interaction with an IR system, the goal state, therefore, is to bring the user to a state of readiness to find, understand and use information for the performance of the task for which the information is being sought. How can we further operationalize this state of readiness?

3.4. Selection State: Information Channel

Before going on, we would like to emphasize a distinction. The goal state we envisage for the user as a result of system-user interaction is not the satisfaction of the information need, principally because the information stimuli and information process that will generate the sought after knowledge cannot be predicted by the user before it occurs (Ford, 2005, Chap. 5, this volume). The goal state is to set the user on the right path or channel to potential information and knowledge. Therefore, the goal state we envisage for the user in interactive IR is not to find, understand and integrate information into modified and new knowledge structures, but as a substitute for information need, which in current IR systems forms the basis for the user's query or a request statement the user types into the search box at the beginning of the interaction. In other words, the goal state as we propose it here represents the function of the information need statement in traditional IR systems.

We will codify this distinction thusly:

> The goal state of user-system interaction is to put the user into a state of readiness to select a channel for finding information that will (or has a good chance of) satisfy his or her information need. The channel still must be followed, the information found, understood and integrated into the user's previously existing knowledge system, creating what we will call an *informational moment* (of some sort).

This distinction allows us now to conceive of the goal state and informational moment, where the user becomes informed by information, into two parts of a complete process. In a certain sense, the goal state is the first part of an information process the end of which is the user becoming informed by the information he or she has been channeled to by the IR system (i.e., from the system output in the results list) (cf. also, Brookes, 1980; Cole, 1994; Todd, 1999).

Because we believe the responsibility for an information process must be assumed by an interactive IR system, our modeling of the goal state of the user due to system-user interaction must also contain the potentiality of the information process to come (we repeat, which the interaction is responsible to provide even though the user may leave the system, or get waylaid and cease the information search/seeking before the satisfaction of his or her information need).

The effect of an information process in the cognitive approach is the transformation of the user's knowledge structure or cognitive state (see above), which was put into a formula by Brookes (1980) in his famous fundamental equation for information science:

$$K[S] + \Delta I = K[S + \Delta S]$$

The equation describes the user's knowledge structure before (KS) and after (K[S +ΔS]) information transforms it (ΔI)—information is defined as that which modifies knowledge structure. Todd (1999) states:

> The [Brookes' fundamental equation] asserts that when people purposefully select information from the multitude of sense impressions that they encounter in their daily lives, and integrate this into their existing private knowledge, it can make a difference to what they already know. It asserts that the process of doing something with selected information has a cognitive effect. A cognitive change takes place. Existing private knowledge, organized as a structured, integrated and coherent entity, is transformed by this information, creating a new knowledge structure (Todd, 1999, p. 864; emphasis added).

It is evident that the ultimate goal of a goal state for a user using an IR system to seek information is for an information process to occur, but not necessarily when using the IR system. The system's role in user-system interaction is to open or establish a channel from the user's goal state for the user to follow to an information process.

We hypothesize that the system can open a channel to an information process by putting the user into a selection state. We highlight the concept of selection because Brookes invoked the term (above), and we have previously defined information in terms of selection:

> Information = the uncertainty associated with the degree of freedom of the act or process whereby one message is chosen or selected from a set of possible messages . . . . (Cole, 1993, p. 209)

Therefore, because of the informational aspect of the channel opening, an interactive IR system can stimulate the opening, putting the user into a selection state by sending a system message to the user with a set of alternative paths to information from which the user must make a selection. We show the channel in Figure 2 with the large arrow

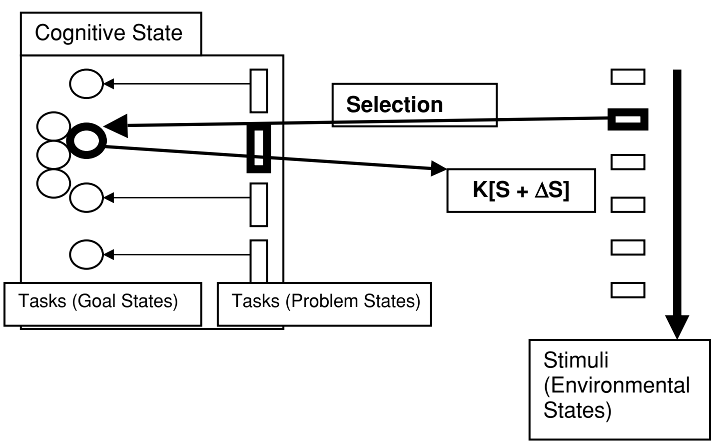

Figure 2: The selection state-information channel model of cognitive IR interaction

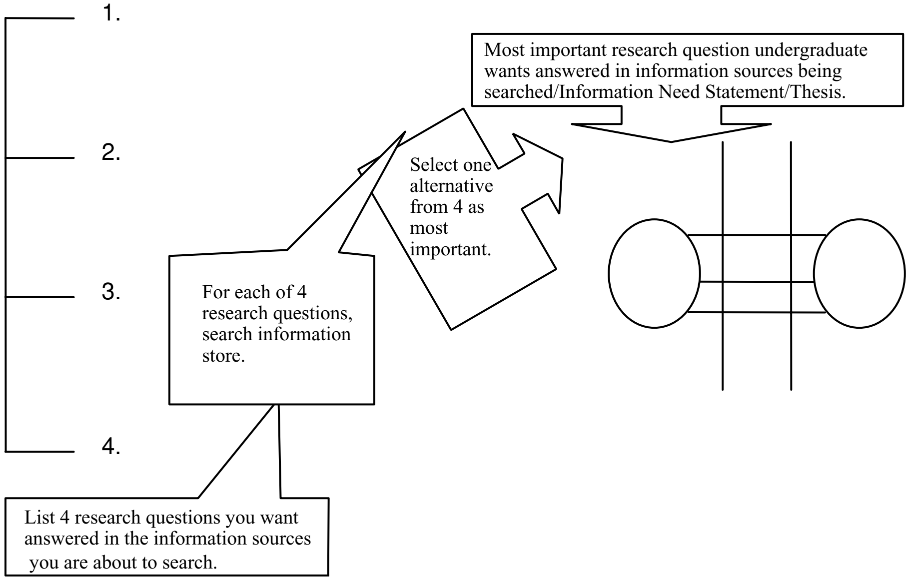

Figure 3: The selection state-information channel model for undergraduates seeking information for a course essay: The system message stimulating the user into a selection state

starting from the system message, shown as passing states of environmental stimuli, stimulating the user's goal state into a selection state, then reversing itself as a channel to data or information which will invoke, if the user finds, understands and integrates it, an information process (indicated by the effect of information in the Brookes' equation).

In Figure 2, the user-system interaction is represented as a series of interacting states: the user's cognitive state, the user's task or problem states, and the various goal states for each task or problem that the user brings to the interaction with the IR system. In turn, the IR system presents the user with information stimuli that we show as passing environmental states. The environmental states can correspond or not with the user's cognitive, task or information need states. It is the responsibility of the user-system interaction to align these states. An example of a system message designed to align the user and system states is given in Figure 3.

When the user selects a system environmental state (i.e., message) that stimulates a goal state/selection state, the system has readied the user to follow a channel to an information process. The alignment of the various states prepares the user for the completion of an information process when the channel to information pointed to by the system is followed, the information found, understood and integrated into a previous knowledge structure of the user, causing transformation of that knowledge structure. This conception of the responsibility of the IR system is not inimical to current IR system design. Part of the focus of IR research, e.g., WebWatcher (Joachims, Freitag, and Mitchell, 1997), is to indicate to the user past selections he or she has made, in order to short cut the user immediately into an appropriate goal state (Kelly, 2005, Chap. 9, this volume; cf. also, Tedeschi, 2005).

3.5. Selection State: Example

We present an example of the Selection State-Information Channel Model taken from a study partially reported in Cole et al. (2005). The next section presents different results from the same study. The study population is undergraduates researching a social science course essay. The start state for the undergraduate is defined as the topic statement for the course essay. The selection state is the selection of the thesis statement for the essay. Here, we make the assumption that the thesis statement is the most appropriate goal state, and that it is at the center of the information need of the undergraduate user—i.e., the user needs information in support of the thesis statement, and because all argumentation in the essay must be in support of the thesis statement the thesis also defines the structure of the essay as well as information need to create the structure. The system message designed to achieve the selection state is diagramed in Figure 3.

From Figure 3, from left to right: The selection state of the undergraduate is achieved by the system message asking the undergraduate to write four research questions, then select the most probable question, which becomes the thesis statement for the essay. The concepts of the thesis statement are then organized into an index structure, (in this case) a compare and contrast essay (the two circles joined by three comparison concepts). The structure of the essay is defined by the thesis statement. The information needed is also governed by the structure as defined by the thesis statement. Therefore, the structure indexes the undergraduate's mental model of the essay (which can be changed), and also creates a query mechanism to the IR system.

4. SPECIFIC PROBLEM

The last part of the chapter reports findings from a study of "selection" carried out in a naturalistic setting of history and psychology courses given at McGill University in 2003. The study tests the hypothesis that a user can, through interaction with an IR system in a single search session, be taken from a start state to a selection state by asking the undergraduate to make a selection of a thesis statement from four possible alternatives.

5. THE STUDY

In the winter term, January to April, 2003, eighty undergraduate students attending five separate history and psychology courses at McGill University were interviewed using the interview schedule shown in Appendix A. McGill University is a large (26,000 full-time students), North American research university offering doctoral programs in a full range of disciplines, including medicine and law. The 80 students were interviewed about their topic for a course essay, an essay that would be evaluated by the course instructor and used as a component of the total mark the student received for that course on the student's transcript. The interviews, therefore, were about the student's real information need.

The students in the five courses represented both the humanities or "soft" end of the social science spectrum (history students) and the science or "hard" end of the social science spectrum (psychology students) (cf. Whitmire, 2002), as well as both inexperienced 1st year students and more experienced 3rd year students.

History 100: Twenty students from one large (over 150 students) 200 Level (1st year) history course.

Psychology 100: Twenty-one students from one large (over 150 students) 200 level (1st year) psychology course.

History 400: Nineteen students from two small courses of nine and ten students each (10–20 students) 400 Level (3rd–4th year) history courses.

Psychology 400: Twenty students from one large (over 100 students) 400 Level (3rd–4th year) psychology course.

5.1. Data Collection

The researchers entered the five courses in the first or second week of classes to gather volunteers for the research project. The sample of students was self-selected and thus does not necessarily represent the larger population of history or psychology undergraduates at McGill University or elsewhere. We then contacted the volunteers by e-mail, asking them to contact us when they had selected their essay topic and had done some preliminary thinking about it (i.e., when they were in Stage 3 of Kuhlthau's (1991) Information Search Process (ISP) Model). The interview schedule required the students to list subject terms or to draw visualizations of their topic on large sheets of newsprint-type paper. The interviews were video/audio taped and lasted from 30 minutes to 1 hour.

The interview schedule's overall aim is to facilitate the undergraduate's narrowing of his or her initial Topic Statement (Question 1 in the interview schedule) to a Selection State Statement (Question 9). The interview schedule facilitates the subject's narrowing of his or her Topic Statement to a Selection State Statement via various cognitive activities:

- listing
  - topic terms (Questions 2 and 2b)
  - four research questions (Question 6)
- visualizing
  - the topic statement (Questions 4 and 5)
  - the most important research question (Question 8)
- ranking
  - concept terms related to their topic (Question 3)
  - the four research questions (Question 7)
  - concept terms related to the Selection State (Question 10).

The study observes the effect of the above listed activities for promoting cognition in the subject. Dependent variables were preference indications by the subjects themselves and counting and comparing the number of concept terms listed by the subject at various points in the interview schedule. The data from the interviews was inputted into SPSS which calculated analysis of variance (ANOVA).

5.2. *Limitations of the Study*

1. The study tries to gauge in a very general sense the cognitive effect of the listing, visualizing and ranking activities for promoting cognition in the subject, using such effect variables as subject preference or by counting and comparing the number of concept terms listed by the subject at various points in the interview schedule. Counting concept terms at different points in a search process, as we do here, or in an information seeking process, as was done by Vakkari, Pennanen, and Serola (2003), is a new, exploratory path for cognitive IR research. In our study, we must make inferences from the quantitative measurements we have made concerning user concept terms so that the findings comment on the specific problem we wish to address in this chapter.
2. We make the assumption that the subject gives a Selection Statement at the end of the interview. In the interview, we did not refer to the students' last statement as the Selection State Statement because this is a technical information science term that would have confused the research subjects.
3. The cognitive activities we asked the subjects to carry out in the interview schedule were always in the same order (for all subjects), which is a limitation of the research instrument.
4. Are findings a function of the course assignment or the cognitive activities given in the interview schedule? The undergraduate courses History 100 and History 400, as well as Psychology 400, all had an essay requirement. For the Psychology 100 course, however, it was impossible to find a course with a pure essay requirement during our three year research project. We decided to interview students from a Psychology 100 course entitled "Social Psychology." The title and the course requirement seemed promising. The students had to envisage a social psychology experiment, with a General Problem statement, a limited Review of the Literature, a Specific Problem statement, as well as an Hypothesis statement. We believed this had the potential to be similar to an essay structure. However, when we analyzed the students' proposals from the point of view of essay structure, we found this course was different from the other courses in the study, with most student essays having a "cause and effect" essay structure. The other courses all had a much greater mixture of essay structure types (Cole et al., in press). Therefore, we took the relevant study findings (see below) and reanalyzed them without the Psychology 100 group (Figures 12, 13, and 14).

5.3. *Findings*

- Students on average dropped 2.6 terms from the original list of terms in their Topic Statements, when writing their Last List (LL) of terms (Table 1). Figure 4 shows the frequency distribution of terms omitted from the last list. Approximately 19 percent did not omit any of the original terms from their final list, and 24 percent dropped only one term. About 34 percent, however,

Table 1: Mean data

Group ID	OL terms omitted	Terms added to last list	OV terms omitted	Terms added to CLV	OV relations omitted	Relations added to CLV
History 100	1.90	2.90	.05	.00	.95	2.85
History 400	3.21	3.63	.16	.05	1.89	3.16
Psych 100	4.14	3.14	.71	.62	1.81	2.86
Psych 400	1.25	2.00	.05	.25	.95	2.15
Total	2.64	2.91	.25	.24	1.40	2.75

omitted more than 8 or more terms, indicating very active cognitive activity. Analysis of variance (ANOVA) shows that the four groups of students in the study differed significantly from each other in terms of number of omitted terms ($F = 8.678$, $(3,76)$, $p < 0.000$). As Figure 5 indicates, students in the first-year psychology course omitted more terms than those in other groups, followed by students in the fourth-year history courses.

- Students on average added 2.9 terms to their final list, which were not in their original list. Figure 6 showing the frequency distribution of terms

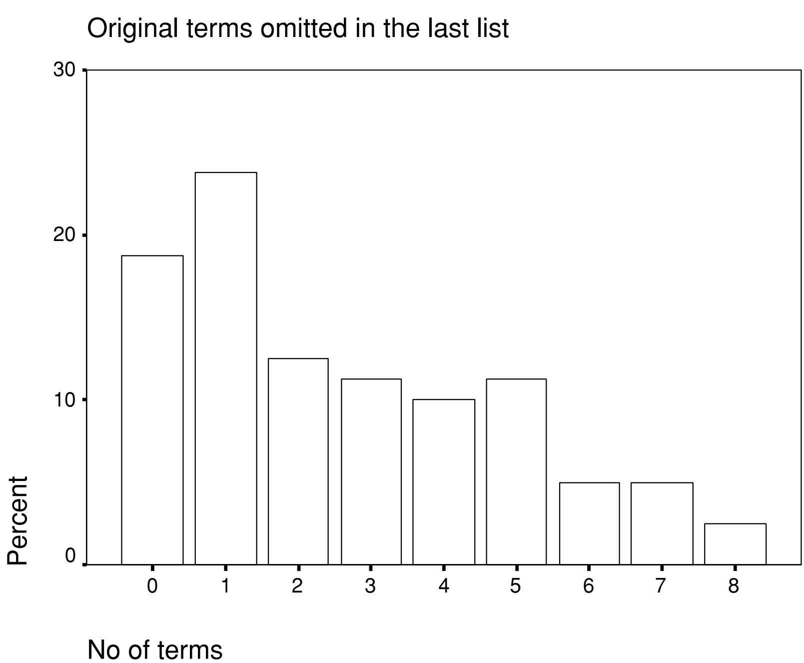

Figure 4: Original terms omitted in the last list

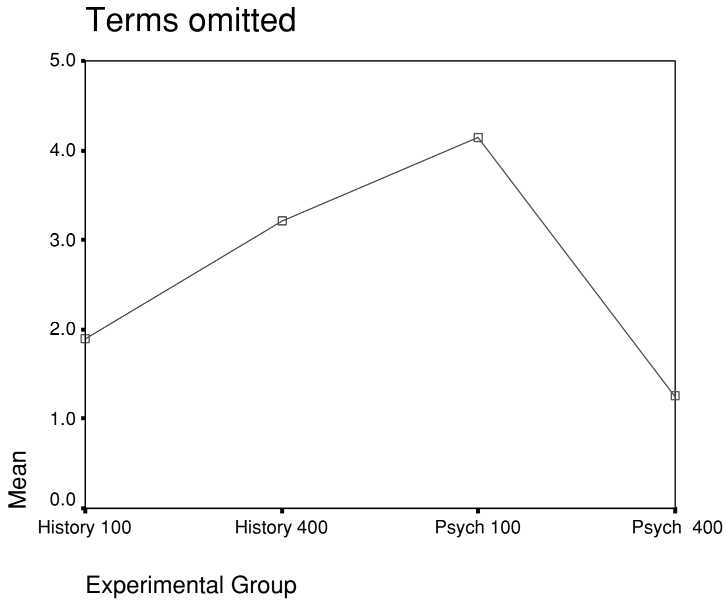

Figure 5: Original terms omitted in the last list

added to the last list, shows that 15 percent did not add any new terms, 19 percent added one term, 14 percent two terms, and 20 percent three terms. Some 20 percent of the students added between 4 and 10 terms to their final list, indicating very active cognitive activity. No significant

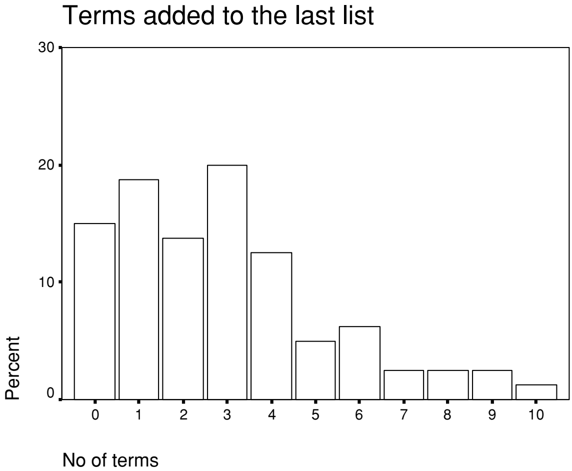

Figure 6: Number of terms added to the last list

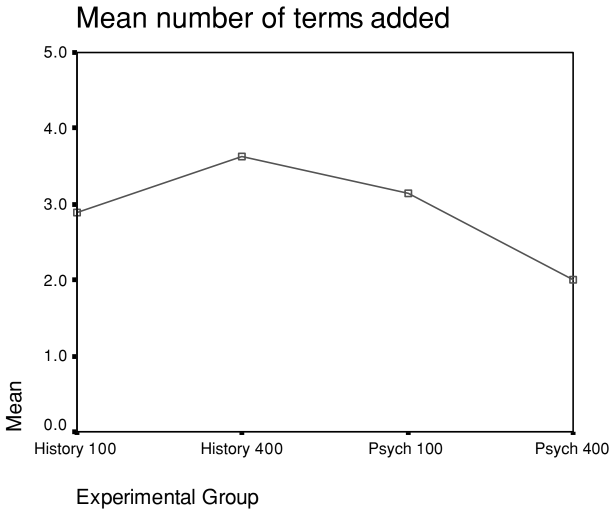

Figure 7: Mean number of terms added to last list per group

results were observed among groups. ($F = 1.646$, (3,76), $p = 0.186$) (Figure 7)

- Students employed a variety of styles in their Original Visualization (OV) when we asked them to visualize their concept terms and the relationships (rel) between terms. While one percent used a bubble style to draw their ideas, slightly more than 26 percent used arrow grams, 24 percent circles and lines, and 16 percent flowcharts. Students also used a variety of styles to visualize the selected question, called the Question Visualization (QV), from the four research questions, including arrow grams (36 percent), a combination of different styles (19 percent), and circles and lines (16 percent) (Figure 8). A significant number of students changed their visualization techniques from their Original Visualization (OV) to their selected Question Visualization (QV) (Chi-square $= 80.010$, $df = 16$, $p < 0.000$).
- Preference was determined by asking students which of the methods used through the interview session would they prefer to use for analyzing their research topics. Slightly more than 32.5 percent preferred to use the last list (LL) of terms that they devised for formulating their queries. This is followed by the 21 percent who preferred circle and line visualization (CLV) method of visualizing their queries, and 18.8 percent who liked to map out the final preferred questions (Figure 9).
- ANOVA indicates that there are no significant differences between history and psychology students for the number of terms omitted or added in the subsequent stages of the interview, and number of relations omitted or added in the circle and line drawings ($p > 0.05$) (Figure 10).

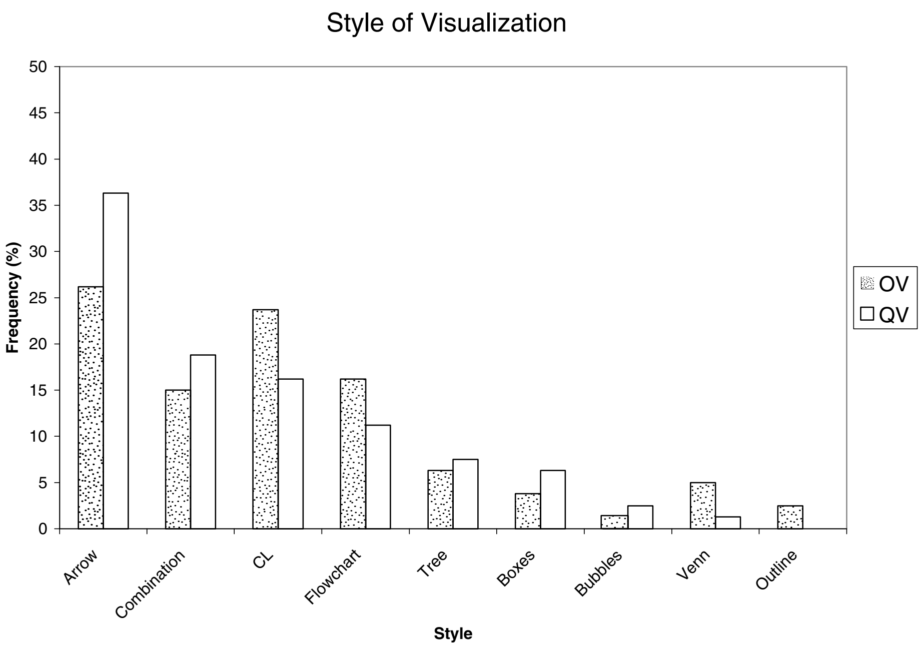

Figure 8: Style of visualization used in their Original Visualization (OV) and for selected Question Visualization (QV)

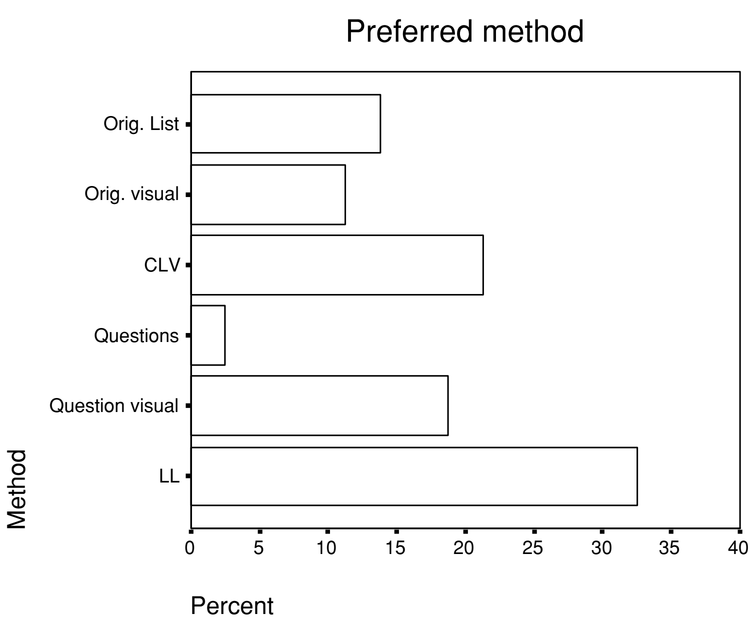

Figure 9: Preferred method for finding information sources

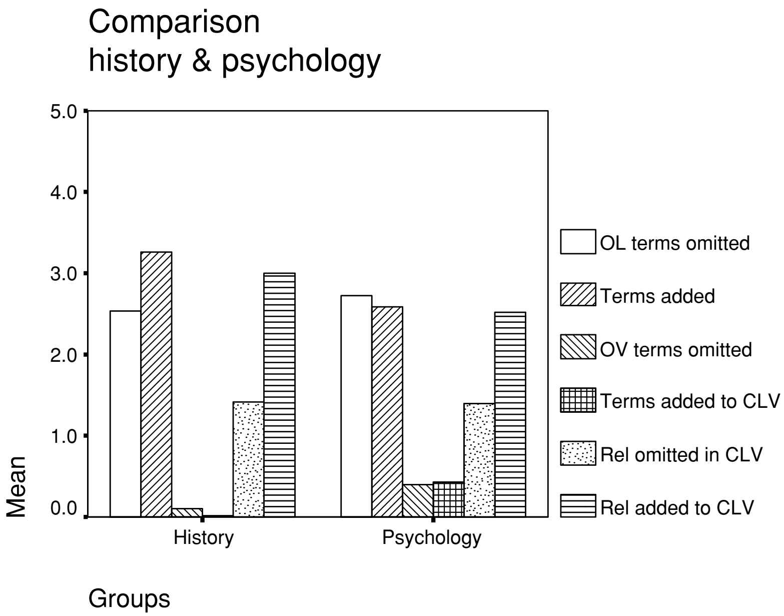

Figure 10: Comparison of history and psychology students for terms added/omitted and relations added/omitted in CLV

- Similarly, there were no significant differences between students in first-year courses and those in higher level courses in terms of number of terms and relations omitted or added ($p > 0.05$) (Figure 11)

Data analysis was also conducted by deleting the first-year psychology course from the analysis. The first-year psychology course was excluded from these analyses due to the nature of their assignment, which was effectively different from the other three groups (see above, Section 5.2 Limitations of the Study), leaving three experimental groups only: students in first-year history course, students in fourth-year history course, and students in fourth-year psychology course.

- As in previous results for all four groups together, a significant difference was observed among the three groups in their omission of terms from their last lists (LL) ($F = 5.679$ (2,56), $p = 0.006$) (Figure 12). Students in the fourth-year history course omitted more terms than the other groups.
- For our three group analysis, two more ANOVA tests are of interest. There seems to be some difference among groups in the number of terms added to their circle and line visualizations (CLV) ($F = 2.895$ (2,56), $p = 0.064$) (Figure 13).
- Another difference among the three experimental groups appears to be in the number of terms added to the last list (LL), which were not in the original list ($F = 2.417$ (2,56), $p = 0.098$) (Figure 14). Students in the fourth-year history course added more terms to their last list than other groups.

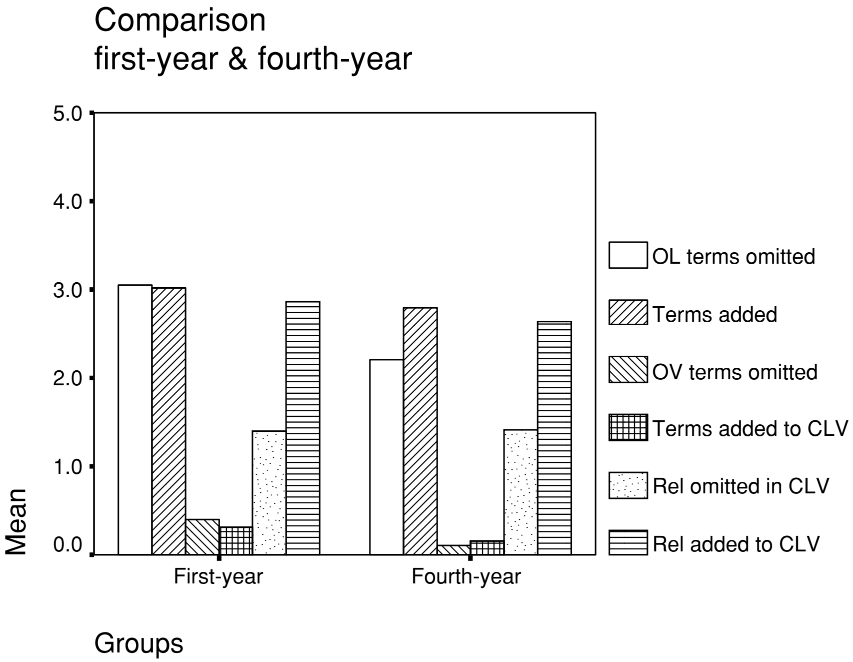

Figure 11: Comparison of first and fourth year students for terms added/omitted and relations added/omitted in CLV

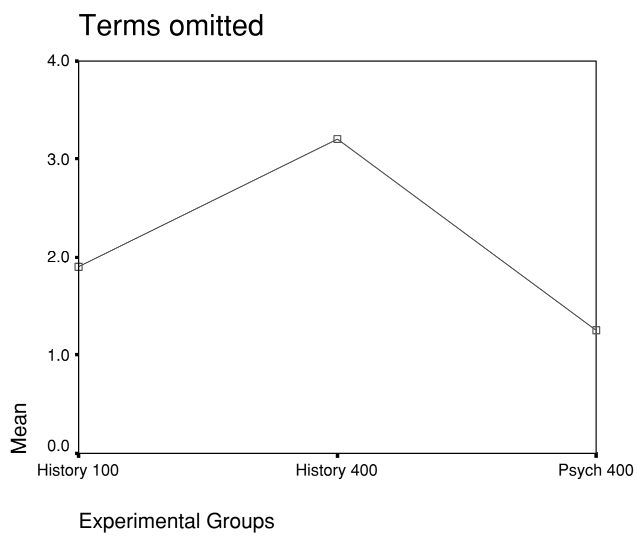

Figure 12: Three group analysis: Terms omitted from LL

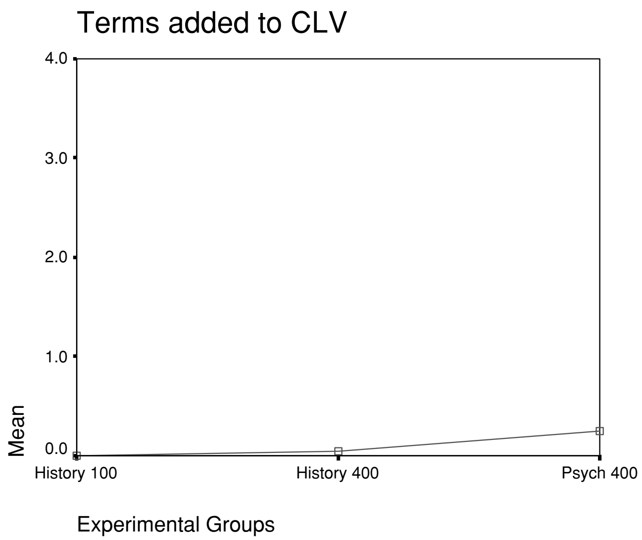

Figure 13: Three group analysis: Terms added in CLV

5.3.1. Findings: highlights

- Statistically significant results: Omitted terms from Original List (OL) of terms to the Last List (LL) of terms differed in a statistically significant amount for both the 4 group comparison (Table 1, Figures 4 and 5) and the 3 group

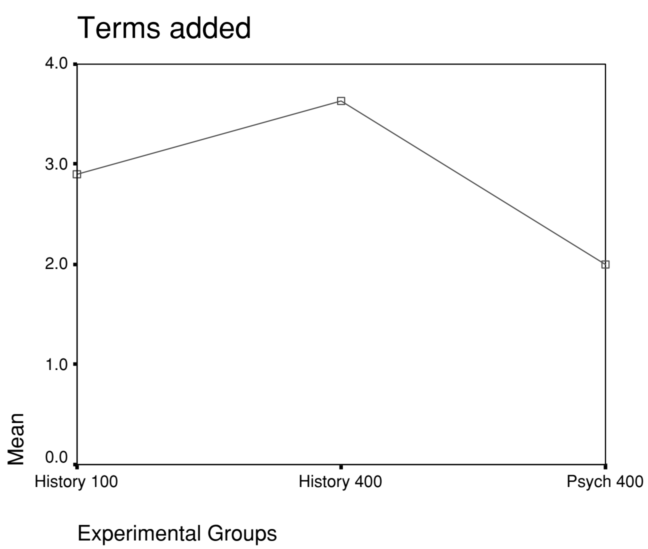

Figure 14: Three group analysis: Terms added in LL

comparison (Figure 12). For the 4 group comparison, as Figure 5 indicates, students in the first-year psychology course omitted more terms than those in other groups, followed by students in the fourth-year history courses. For the 3 group comparison, as Figure 12 indicates, students in the fourth-year history courses omitted more terms than those in other groups. Also, a statistically significant number of students switched visualization styles from OV to QV after they had selected the research question most important to their essay (Figure 8).

- In general, students in the fourth-year history course show more active cognitive activity than students in the other experimental groups. They tended to omit more terms from their original list in their final list of query terms, and replaced these with new terms. Similarly, they seem to have omitted more relations among terms from their original visualization drawings in their circles and lines diagrams, and replaced these with new relations.
- Students in the first-year psychology course, on the other hand, showed slightly less cognitive activity than students in the fourth-year history course. Again with this group we observe a similar pattern: they tended to omit more terms from their original list and replaced these with new terms, and they seem to have omitted more relations among terms from their original visualization diagrams and replaced these with new relations.
- Discipline and course-level did not have any effect on cognitive styles chosen by students.
- Visualization did not result in students adding or omitting terms for their lists.
- Visualization styles used most by students were arrow grams, then in this order: a combination of different methods, circle and lines, and flow charts.
- The preferred method for accessing information from information sources was the last list (LL) of concept terms complied by the students in the interviews, followed by circle and line visualization, and the four questions visualization.

5.4. Discussion

Popular search engines interfaces, as well as other information accessing interfaces (e.g., to scholarly databases like PsycINFO, ERIC, etc.), require the user to either type in key words or to specify in some other way their query to the system, thus creating a word-representation agent of their information need. For undergraduates researching their course essay, the study's findings tentatively support this chapter's premise that the undergraduate's topic statement, given at the beginning of their interview with us, is not their information need as the vast majority of study subjects indicated one of the later stages of the interview schedule as their preferred approach to accessing information for their essay. Slightly more than 32.5 percent preferred to use the last list (LL) of terms for formulating their queries to information sources.

We made the assumption that undergraduates' last list (LL) of terms in the interview was equivalent to our theoretical notion of the user's Selection State in our Selection State-Information Channel Model. In the Model, the goal state for the user of an IR system is to be in a Selection State, which occurs after the user is stimulated to make a selection from a set of alternatives. This is so because making a selection changes the cognitive state of the user, setting up an appropriate channel to information and an information process. In the study, the undergraduates were asked to list 4 research questions they wanted answered by information sources, then to select the one research question that was most important to their essay. Further research is needed to test if asking the subject to list four research questions then to make a selection from the list actually puts the study participant in a Selection State.

The results tentatively indicate that the Selection State Statement, called the Last List (LL), and given by the subjects at the end of the interview, may be a different conception or perspective from the Topic Statement, called the Original List (OL), given by the subjects at the beginning of the interview schedule. A statistically significant result was that the subjects switched style of visualization from the Original Visualization (OV) and the Question Visualization (QV) given after the students were asked to list four research questions and select the most important one for their essay. In both visualizations, the subject could choose whatever style he or she wished. However, other factors besides a different cognitive state between the OL and OV and the LL and QV may have intervened, such as the intervening interview question asking the subject to use circles and lines to indicate relationships between the lines.

(The evolution of concept terms over the course of some activity is an exciting new area of information science research. For a research study measuring the change in the number of student concept terms over a longer period of time (three months between measurements), see Pennanen and Vakkari, 2003. For an example of research that also quantifies change in users' knowledge structure (according to uncertainty fluctuations), but over Kuhlthau's (1991) six-stage ISP model, cf. Wilson et al., 2002.)

6. CONCLUSIONS AND FUTURE DIRECTIONS

We argue that many undergraduates researching a course essay approach an IR system when they are in ISP Stage 3, an exploration stage of researching an assignment with an unknown item information need—i.e., they do not know and are not able to specify their information need in a query to the system. In order to upgrade performance and user satisfaction, new interactive IR systems must go beyond their present reliance on this type of user typing in topic concepts as a representation of their information need.

This chapter started from the cognitive approach's dissatisfaction with information need as a viable agent for the user in the user-system interaction. We have in this chapter attempted to reconceptualize the model of user-system interaction, with its assumption of a static, constant information need, so that the user's agent in the interaction better mirrors the actual information need of the user. To do with this, we have conceived of the user's agent as "in-for-mation" at the beginning of the interaction. The purpose of the interaction is to stimulate the user into the formation of a goal state for the interaction, which is to put the user into a Selection State. That is, the system interaction must ready the user for an information process, the end part of which occurs when the user actually finds, understands and integrates the information pointed to by the system.

Given that the purpose of the user-system interaction is to ready the user for an information process, the Selection State-Information Channel Model indicates a channel from the user's Selection State to information which, if found, understood and integrated into the user's knowledge structure, will produce an information process. The Selection State-Information Channel Model reconceptualizes the role of the IR system.

The new role of the IR system is to make the user's reason for seeking information the first priority of the system. The user does not use an IR system to find information in of itself, but rather to find information that will facilitate the performance of some task. This can be a work task or an everyday life task. Putting the user's motivation for seeking information up front in the interaction model, sets certain priorities for new directions in the research and design of IR systems. We believe that a thorough investigation of the nexus between the user's task and the role of information in the performance of that task should be our first priority.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of the Social Sciences and Humanities Research Council (SSHRC) of Canada, File No. 410-2002-1394.

APPENDIX

Interview Schedule: January–April 2003

1st sheet	1.	Could you please describe and write down the topic of your paper?
	2.	What words, search terms, keywords or concepts will you use to look for information to write your paper? Please write out the terms.
	2b.	Include any terms from the topic sentence.
	3.	Could you please label these terms, using initials or abbreviations? Could you rank these terms according to their importance in your paper?
2nd sheet	4.	Could you in some way visualize, draw, map or chart these words, terms or concepts in your paper showing the relationship(s) between them, if any? *Use the labels you have chosen. [When finished]: Please write the search term(s) represented by each label on the diagram. Please verbally explain your diagram.
3rd sheet	5.	Could you now visualize these same terms, using circles and lines this time? The circles should vary in size according to importance of terms. Their distance apart should be representative of how closely related the concepts are. The lines should show inter-relationships amongst the terms. *Please use the labels you chose for each term to label the circles rather than write the term in the circle. [When finished]: Please write the search term words on the diagram and please verbally explain your diagram.
4th sheet	6.	Could you please now write down 4 questions you would like answered by your information source?
	7.	Could you please rank these 4 questions by their importance to your paper?

5th sheet	8.	Concerning the question you just ranked as most important, please write out the terms you would use to look for information for that question. Please label these terms with initials or abbreviations. Could you map out these terms in a new diagram? Please use these labels on your diagram. [When finished]: Please write the search term words on the diagram and please verbally explain your diagram.
6th sheet	9.	Having gone through this process, can you now write down what your paper will be about? Feel free to make changes or integrate new thoughts.
	10.	What words, search terms, keywords or concepts would you use now? Please write them down, and rank these terms according to their importance to your paper.
	11.	Of these approaches [the original list, your initial visualization, the circles and lines diagram, the last diagram, or your final list of terms] which would you find most useful in finding information to write your paper? Please mark the one you prefer and verbally explain why would you prefer it.

REFERENCES

Bates, M. J. (1990). Where Should the Person Stop and the Information Search Interface Start? *Information Processing and Management*, 26(5), 575–801.

Beaulieu, M. (2000). Interaction in Information Searching and Retrieval. *Journal of Documentation*, 56(4), 431–439.

Beheshti, B., Bowler, L., Large, A., and Nesset, V. (2005). Towards an Alternative Information Retrieval System for Children. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 139–165). Dordrecht: Springer.

Belkin, N. J. (1980). Anomalous States of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science, 5*, 133–134.

Belkin, N. J. (1980). Anomalous States of Knowledge as Basis for Information Retrieval. *The Canadian Journal of Information Science*, 5, 133–143.

Belkin, N. J., Cool, C., Stein, A., and Thiel, U. (1995). Cases, Scripts, and Information-Seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications*, 9(3), 379–395.

Belkin, N. J., Oddy, R., and Brooks, H. (1982). ASK for Information Retrieval. *Journal of Documentation*, 38(2), 61–71.

Bopp, R. E. and Smith, L. C. (2001). *Reference and Information Services: An Introduction*. Englewood, CO: Libraries Unlimited.

Brookes, B. C. (1980). The Foundations of Information Science. Part I. Philosophical Aspects. *Journal of Information Science*, 2, 125–133.

Cole, C. (1993). Shannon Revisited: Information in Terms of Uncertainty. *Journal of the American Society for Information Science*, 44(4), 204–211.

Cole, C. (1994). Operationalizing the Notion of Information as a Subjective Construct. *Journal of the American Society for Information Science*, 45(7), 465–476.

Cole, C., Kennedy, L., and Carter, S. (1996). The Optimization of Online Searches through the Labelling of a Dynamic, Situation-dependent Information Need: The Reference Interview and Online Searching for Undergraduates Doing a Social Science Assignment. *Information Processing and Management*, 32, 709–717.

Cole, C., Leide, J. E., Large, A., Beheshti, J., and Brooks, M. (2005). Putting it Together Online: Information Need Identification for the Domain Novice User. *Journal of the American Society for Information Science and Technology*, 56(7), 684–694.

Cole, C., Leide, J. E., Beheshti, J., Large, A., and Brooks, M. (in press). Investigating the ASK Hypothesis in a Real-life Problem Situation: A Study of History and Psychology Undergraduates Seeking Information for a Course Essay. *Journal of the American Society for Information Science and Technology*.

Cutter, C. A. (1876/1904). *Rules for a Dictionary Catalog* (3rd Edition). Washington: Government Printing Office.

Cutter, C. A. (1891–1893). *Expansive Classification: Part 1: The First Six Classifications*. Boston, MA: Cutter.

Efthimiadis, E. N. and Robertson, S. E. (1989). Feedback and Interaction in Information Retrieval. In C. Oppenheim (Ed.). *Perspectives in Information Management* (pp. 257–272). London: Butterworths.

Fischer, G., Henninger, S., and Redmiles, D. (1991). Intertwining Query Construction and Relevance Evaluation. In *CHI'91 Proceedings of the Conference on Human Factors in Computing Systems*, New Orleans, Louisiana, April 27–May 2, 1991 (pp. 55–62). New York: ACM Press.

Ford, N. (2005). New Cognitive Directions. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 81–96). Dordrecht: Springer.

Herken, R. (1995) (Ed.). *The Universal Turing Machine: A Half-Century Survey*. Wien: Springer-Verlag.

Hearst, M. A. (1999). User Interfaces and Visualization. In Ricardo Baeza-Yates and Berthier Ribeiro-Neto (Eds.). *Modern Information Retrieval* (pp. 257–323). New York: ACM Press.

Hook, P. A. and Borner, K. (2005). Educational Knowledge Domain Visualizations: Tools to Navigate, Understand, and Internalize the Structure of Scholarly Knowledge and Expertise. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 187–208). Dordrecht: Springer.

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.

Jansen, B. J., Spink, A., and Pedersen, J. (2005). A Temporal Comparison of AltVista Web Searching. *Journal of the American Society for Information Science and Technology*, 56(6), 580–570.

Joachims, T. Freitag, D., and Mitchell, T. (1997). WebWatcher: A Tour Guide for the World Wide Web. In *Proceedings of the International Joint Conference on Artificial Intelligence (ICML 97)*, USA, 770–777.

Katz, W. A. (1997). *Introduction to Reference Work: Reference Services and Reference Processes*, Vol. 2. New York: McGraw-Hill.

Kelly, D. (2005). Implicit Feedback: Using Behavior to Infer Relevance. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 169–186). Dordrecht: Springer.

Kitajima, M. and Polson, P. G. (1996). A Comprehension-based Model of Exploration. In *CHI'96 Proceedings of the Conference on Human Factors in Computing Systems*, Vancouver, BC, April 13–18, 1996 (pp. 325–331). New York: ACM Press.

Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.

Lalmas, M. and Ruthven, I. (1999). A Framework for Investigating the Interaction in Information Retrieval. In *Proceedings of the 9th European-Japanese Conference on Information Modeling and Knowledge Bases* (pp. 222–239). Iwate, Japan.

Larsen, B. and Ingwersen, P. (2005). Cognitive Overlaps Along the Polyrepresentation Continuum. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 43–60). Dordrecht: Springer.

Leide, J. E., Large, A., Beheshti, J., Brooks, M., and Cole, C. (2003). Visualization Schemes for Domain Novices Exploring a Topic Space: The Navigation Classification Scheme. *Information Processing and Management*, 39(6), 923–940.

Lucas, W. and Topi, H. (2002). Form and Function: The Impact of Query Term and Operator Usage on Web Search Results. *Journal of the American Society for Information Science and Technology*, 53(2), 95–108.

Lucas, W. and Topi, H. (2005). Learning and Training to Search. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 209–226). Dordrecht: Springer.

MacMullin, S. E. and Taylor, R. S. (1984). Problem Dimensions and Information Traits. *The Information Society*, 3(1), 91–111.

Markoff, J. (2005). Your Internet Search Results, in the Round. *New York Times*, Monday May 9, 2005, p. C3.

Moran, T. P. (1981). Command Language Grammar: A Representation Scheme for the User Interface of Interactive Systems. *International Journal of Man-Machine Studies*, 15(1), 3–50.

Nathan, M. J. (1990). Empowering the Student: Prospects for an Unintelligent Tutoring System. In *CHI'90 Proceedings of the Conference on Human Factors in Computing Systems*, Seattle, Washington, April 1–5, 1990 (pp. 407–414). New York: ACM Press.

O'Day, V. L. and Jeffries, R. (1993). Orienteering in an Information Landscape: How Information Seekers Get from Here to There. In *INTERCHI'93 Proceedings of the Conference on Human Factors in Computing Systems, Amsterdam, The Netherlands*, 24–29 April 1993 (pp. 438–445). New York: ACM Press.

Oxford Dictionary of Current English (1984). Oxford: Oxford University Press.

Pomerantz, J. (2005). A Linguistic Analysis of Question Taxonomies. *Journal of the American Society for Information Science and Technology*, 56(7), 715–728.

Russell, D. M., Stefik, M. J., Pirolli, P., and Card, S. K. (1993). The Cost Structure of Sensemaking. In *INTERCHI'93 Proceedings of the Conference on Human Factors in Computing Systems*. Amsterdam, The Netherlands, 24–29 April 1993 (pp. 269–276). New York: ACM Press.

Ruthven, I. (2005). Integrating Approaches to Relevance. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 61–80). Dordrecht: Springer.

Saracevic, T. (1996). Interactive Models in Information Retrieval (IR): A Review and Proposal. In *Proceedings of the 80th Annual Meeting of the American Society of Information Science*, 33, 3–9.

Shannon, C. E. (1949). The Mathematical Theory of Communication. In C. E. Shannon and W. Weaver (Authors). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Spink, A. (1997). Information science: A Third Feedback Framework. *Journal of the American Society for Information Science*, 48(8), 728–740.

Spink, A. and Cole, C. (2005). A Multitasking Framework for Cognitive Information Retrieval. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 99–112). Dordrecht: Springer.

Spink, A. and Saracevic, T. (1997). Interaction in Information Retrieval: Selection and Effectiveness of Search Terms. *Journal of the American Society for Information Science*, 48(8), 741–761

Spink, A., Wolfram, D., Jansen, B.J., and Saracevic T. (2001). Searching the Web: The Public and their Queries. *Journal of the American Society for Information Science and Technology*, 52(3), 226–234.

Storrs, G. (1994). A Conceptualization of Multiparty Interaction. *Interaction with Computer*s, 6(2), 173–189.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press.

Taylor, R. S. (1968). Question-negotiation and Information Seeking in Libraries. *College and Research Libraries*, 29, 178–194.

Tedeschi, B. (2005, May 23). With Its Home Page, Google Could Get a Bit Closer to Its Users. *The New York Times*, p. C7.

Terwilliger, R. B. and Polson, P. G. (1997). Relationships between Users' and Interfaces' Task Representations. In *CHI'97 Proceedings of the Conference on Human Factors in Computing Systems*, Atlanta, GA, 22–27 March, 1997 (pp. 99–106). New York: ACM Press.

Todd, R. J. (1999).Back to our Beginnings: Information Utilization, Bertram Brookes and the Fundamental Equation of Information Science. *Information Processing and Management*, 35, 851–870.

Vakkari, P. (1999). Task Complexity, Problem Structure and Information Actions: Integrating Studies on Information Seeking and Retrieval. *Information Processing and Management*, 35, 819–837.

Vakkari, P. and Javelin, K. (2005). Explanation in Information Seeking and Retrieval. In A. Spink and C. Cole (Eds.). *New Directions in Cognitive Information Retrieval* (pp. 113–138). Dordrecht: Springer.

Vakkari, P., Pennanen, M., and Serola, S. (2003). Changes of Search Terms and Tactics While Writing a Research Proposal: A Longitudinal Case Study. *Information Processing and Management*, 39(3), 445–463.

Wersig, G. (1971). *Information-Kommuikation-Dokumentation*. Pullach bei Munchen: Verlag Documentation.

Wersig, G. (1979). The Problematic Situation as a Basic Concept of Information Science in the Framework of the Social Sciences—A Reply to N. Belkin. In *New Trends in Informatics and its Terminology* (pp. 48–57). Moscow: VINITI.

Whitmire, E. (2002). Disciplinary Differences and Undergraduates' Information Seeking Behavior. *Journal of the American Society for Information Science and Technology*, 53(8), 631–638.

Wilson, T. D. (1999). Models in Information Behaviour Research. *Journal of Documentation*, 55, 249–270.

Wilson, T. D. (2000). Human Information Behaviour. *Informing Science*, 3(2), 49–56.

Wilson, T. D., Ford, N., Ellis, D., Foster, A., and Spink, A. (2002). Information Seeking and Mediating Searching. Part 2. Uncertainty and its Correlates. *Journal of the American Society for Information Science and Technology*, 53(9), 704–715.

CHAPTER 3

COGNITIVE OVERLAPS ALONG THE POLYREPRESENTATION CONTINUUM

BIRGER LARSEN AND PETER INGWERSEN
Royal School of Library and Information Science, Denmark

1. INTRODUCTION

The principle of polyrepresentation, proposed more than 10 years ago, provides a holistic and explicitly cognitive framework for understanding the processes involved in Information Retrieval (IR) (Ingwersen, 1994, 1996, 2001). While readily applicable to the phenomena encountered in mainstream algorithmic IR research, the main strength of the principle is that it can also be applied simultaneously to the cognitive space of the user—thus integrating the two perspectives into one coherent cognitive framework. The main idea in the principle is that document overlaps generated from representations of different cognitive and functional origins can improve performance in IR systems. This kind of overlaps we entitle "cognitive overlaps".

This chapter outlines the principle of polyrepresentation with a focus on the representations involved. The potentials and problems of the principle are discussed in the light of recent empirical studies, and challenges and opportunities for future research are identified along a polyrepresentation continuum.

2. THE PRINCIPLE OF POLYREPRESENTATION—A HOLISTIC APPROACH TO IR

The principle of polyrepresentation originates in work on establishing a cognitive theory for interactive IR (Ingwersen, 1992). It can be regarded as a result of an effort to demonstrate the *applicability* of this theory (Borlund, 2000). The principle of polyrepresentation was developed during the 1990s. The principle is fully expanded in the *Journal of Documentation* article from 1996, which remains the main publication on the principle (Ingwersen, 1996). Prior to that, the idea of polyrepresentation is mentioned throughout Ingwersen's book *Information Retrieval Interaction* from 1992 as a high precision tool, and an early version was presented at a SIGIR conference (Ingwersen, 1994).

A recent update on the principle appeared at the CoLIS4 conference (Ingwersen, 2002). This section briefly examines the cognitive foundation of the principle of polyrepresentation, outlines its main hypothesis, and gives examples of the kind of representations that can be captured from the cognitive space of the user and the information space of the IR systems respectively.

The cognitive foundation of the principle of polyrepresentation is very apparent in that it is assumed that every agent in IR contributes with their cognitive perceptions

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 43–60.

and interpretations of a given document as seen from their own context. Based on the cognitive viewpoint in Information Science (De Mey, 1980) such interpretations are viewed as the result of transformations of cognitive models or knowledge structures of the agents involved. When recorded they are regarded as representations of the documents. Obviously, the same document may be interpreted by many different agents, which can result in several representations of the same document, representations that have *different cognitive* origins.

For instance, an academic article (itself a representation of the author's knowledge structures at a given point in time) may be represented by controlled and uncontrolled indexing terms assigned by an indexer, as well as subsequently cited by various authors in different contexts. In addition, a number of representations may derive from the same agent but be of different *functional* nature, e.g., the article title, abstract, table captions, references etc. all generated by the author of a text. These two types of representations are normally associated with subject access to the content or aboutness of the documents. Supplementary representations may be generated based on the actions of *selectors* in relation to the isness of the documents (Ingwersen, 2002). These selectors are agents that are responsible for the availability and accessibility of documents, e.g., journal boards, reviewers, and employers etc. who bestow cognitive authority by, e.g., allowing a paper to appear in a given journal.

The cognitive foundation is also apparent in the way the searchers' situations are considered. An information need is not seen as a static entity in its own right but rather as part of a causal structure in which a work or daily-life task to be solved plays an essential role. Based on an analysis of the user's mental structures in relation to the information need it is proposed to extract a number of representations related to the cognitive space of the user. This is discussed in some detail in Section 3 below.

We therefore define a representation as any tangible and recordable entity that has occurred as the result of a transformation of knowledge structures of a cognitive agent. As the number of representations with different cognitive and functional origins rises it is evident that quite a lot of uncertainties and inconsistencies will arise in the representation of documents and information needs. Seen from a cognitive viewpoint this is inescapable, but not necessarily a disadvantage when exploited constructively.

One may view all the communication processes in IR as consisting of interchanges that take place at the *sign level*, i.e., at a linguistic surface level (Smeaton, 1992). When humans are part of IR activities the communication between generators and recipients of information may in addition take place at a *cognitive level*. Thereby the knowledge structures of the human recipient could potentially be affected and modified, and the searcher thus receives information to help solve her work or daily-life task. Because we may only communicate via signs, the information sent by a generator will always be subjected to a cognitive "free fall", and has to be re-interpreted by a human recipient to achieve communication at the cognitive level (Ingwersen, 1996, p. 6). This essential act of interpretation means that uncertainties and unpredictabilities become inherent features of *any* representation in IR. Much of the research and development work, e.g., in Library and Information Science, has been done to reduce such uncertainties and unpredictabilities. Controlled vocabularies like thesauri or extensive cataloguing rules are constructed to create more uniform document representations, as well as to allow users to reduce their uncertainties when formulating their needs.

The principle of polyrepresentation takes a holistic cognitive view by focussing simultaneously on the cognitive structures, and the representations that may be generated from these, both in the cognitive space of the user *and* in the information space of the IR systems. The principle of polyrepresentation is based on the following hypothesis:

> As all tangible representations in IR are the result of interpretations by the involved cognitive actors inconsistency and uncertainty are inherent in the representations. Therefore, if representations of different cognitive and functional origin point to a set of information objects in a particular seeking context, this cognitive overlap presents more relevant information objects than each independent representation. Further, the more different the cognitive representations producing an overlap are in time and by cognitive or functional type, the higher the probability that such information objects are relevant. (Based on Ingwersen, 1996).

The principle of polyrepresentation thus represents an attempt to view the uncertainties and unpredictability as favourable to IR, and to exploit these actively rather than attempting to eliminate them. Inspired by, e.g., Sparck Jones (1990), polyrepresentation entails working through *intentional redundancy*, that is, to represent documents (or information needs) in multiple, complementary ways. We call this redundancy "intentional", since general and non-estimated redundancy in relation to representations of documents may not always be productive in IR. Rather the degree, mode and function of redundancy should be determined by knowledge of the current user's information behaviour. This means that special emphasis should be placed on documents in searcher-dependent overlaps created by intersecting representations of different cognitive and functional origins—cognitive overlaps (See also Figure 2 below).

The purpose of the principle of polyrepresentation is thus to facilitate the exploitation of a multitude of both cognitive and functional representations with a focus on exploiting *different* functional representations from the same agent, as well as on combining representations from *different* cognitive agents in a structured framework. The following two sections examine the cognitive structures available in the cognitive space of the user and the information space of the system.

3. POLYREPRESENTATION OF THE USER'S COGNITIVE SPACE

Early investigations into the formation of the information need have inspired the application of the principle of polyrepresentation on the cognitive space of the searcher. These investigations clearly demonstrate that there are different underlying cognitive reasons behind the development of an information need. This need is seen as the result of communication, sensing or thinking processes, which result in the realisation that something is missing for the solving of a problematic situation. This conception is inspired by the work of, e.g., Mackey (1960), Belkin (1980), Belkin, Oddy, and Brooks (1982), and Dervin and Nilan (1986).

Four such cognitive structures influencing the user are shown in a bottom-up order of causality on the right hand side of Figure 1 below: a work (or daily-life) task/interest; a current cognitive state; a problem space, including a state of uncertainty; and an information need (Ingwersen, 1996).

The four cognitive structures are the result of an analysis of the stability of the user's cognitive-emotional structures in relation to the information need, and places that need in a causal structure: The work task/interest perception, which is influenced

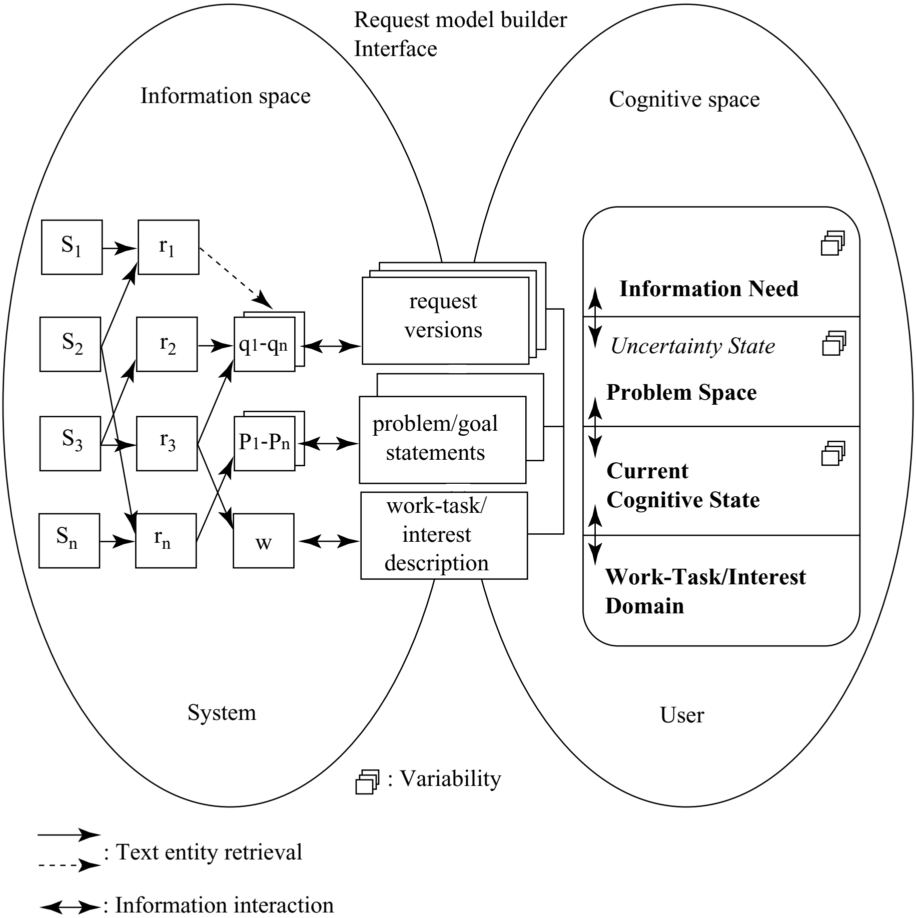

Figure 1: The global model of polyrepresentation. (From Ingwersen, 1996, Figure 8, p. 37)

by a common, professional or scientific domain and dominated by the individual's intentionality and cognitive-emotional state, causes a problematic situation, and, if this cannot be solved without external information, an information need. The current cognitive state includes the (limited) awareness or sense of what is desired. The current cognitive state, the problem space and the information need are all *variable*.

These three structures are easily affected by external input and/or thinking processes, while the work task/interest perception is set in a social context and may be more stable. For example, it may be related to a person's actual work situation as an engineer, to a person's continuing interest in certain aspects of Tolkien's artificial languages or other daily-life situations. The problem space corresponds to Belkin's ASK (Belkin, 1980; 1984) and is separated from the information need because the same problematic situation may give rise to several different information needs.

Based on earlier research it can be inferred that such cognitive-emotional structures have been manifested, and that some of them appear in quite well-defined forms (Ingwersen, 1996). However, as the four cognitive structures reside in the cognitive space of the user we do not have direct access to them during IR, and we have to work with *tangible representations* of them. Such representations or evidence could be extracted by human intermediaries or by interface functionalities during information interaction (indicated as a request model builder (RMB) interface in middle of Figure 1).

Polyrepresentation of the user's cognitive space may thus be achieved by capturing a number of different functional representations from the information seeker, as indicated in the middle of Figure 1. In an ideal situation, potentially at least three *functionally different* representations may be extracted at any one point in time (Ingwersen, 1996):

1. a 'what', i.e., a request version which includes what is currently known about the unknown (the wish or desire for information)
2. the 'why', i.e., a problem statement as well as
3. a work task or daily-life interest and domain description.

Because the underlying cognitive structures are variable over a session, different versions of each representation may occur over time (this temporal variability is indicated in Figure 1 as an increasing number of 'boxes'). Some of the captured representations may often appear to be similar, e.g., the problem statement and the work task description. This is a consequence of the fact that information needs may be well or ill defined, as well as more or less stable.

Evidence of emotional aspects, e.g., of uncertainty, doubt, satisfaction or relief, may be included in statements (Kuhlthau, 1991; 1993). These different types of information needs and their development are clearly demonstrated by the empirical studies (Ingwersen, 1982; Belkin, 1984), as is the role of the librarian in helping the user to define and refine her need. Such a set of representations extracted from the seeker's cognitive space provides a more fertile and *operational context* than a request version alone. The intention in the principle of polyrepresentation is that this enriched set of representations should be combined with each other and used as search terms during interactive IR, in order to achieve polyrepresentation of the information space.

4. POLYREPRESENTATION OF DOCUMENTS IN THE INFORMATION SPACE OF IR SYSTEMS

The information space of IR systems consists of two major components: the documents and the IT components that give access to the documents. The possible representations one may generate from each component are considered below. The *documents* are influenced by several different cognitive agents. Figure 2 below shows an overview of possible cognitively and functionally different representations of the documents. In addition, the figure may be seen as illustrating the sets of documents, and various cognitive overlaps between them that may be retrieved in relation to a request using each of the representations (Ingwersen, 2002).

The contents of the documents reflect the cognitive structures of the *author(s)* in the form of signs, i.e., the transformations of the interpretations, ideas, and cognitive-emotional structures of the authors(s) with respect to their goals and intentionality. Depending on domain, media, and style, a range of functional representations can be

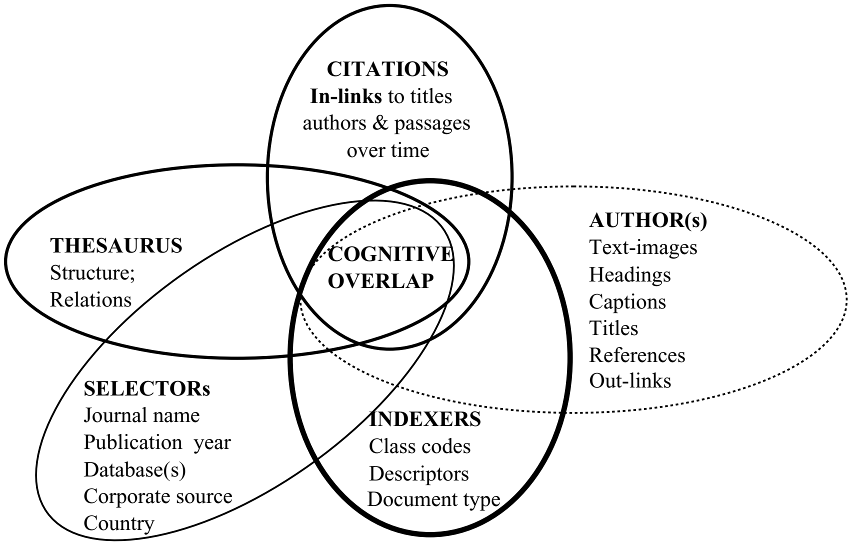

Figure 2: Polyrepresentative cognitive overlaps generated from cognitively and functionally different representations of documents. Retrieved sets are generated by one retrieval engine and associated with one searcher statement (from Ingwersen, 2002, Figure 1, p. 294).

identified in a single document. The representations in Figure 2 are related to academic full text documents—a particularly rich source for generating representations and hence ideal for experiments involving the principle of polyrepresentation.

The functional representations originating from the author that may be captured from the full text of the documents are shown on the right side of the figure. Many representations with strong functional characteristics are available because of the rhetorical structure of the academic documents, commonly organised in specific ways according to convention, e.g., introduction, theory, or methodological sections, results, discussion, and/or conclusions. Like presentation style, the *structural organisation* is domain and media-dependent and very useful as a *supplement* to subject matter. Aside from the structure of the documents, the section titles at different levels and the table and figure captions are examples of functionally different ways of representing a document. These have previously been applied for document representation (Wormell, 1981).

In addition, the *references* in the bibliographies may also be extracted from full text academic documents. The selection of particular references in a given document is seen as highly reflective of the situational factors that affect the author and her current cognitive state at the time of authorship. Similarly, they may signal a kind of situational appropriateness to a potential user who, in her particular situation, might agree to (parts of) the selection of references and thus find the document relevant. An additional opportunity, offered by academic documents in full text, is to identify the text surrounding the location in ci<u>ting</u> documents (if any) where the ci<u>ted</u> document is mentioned, and to use this as a representation of the cited document (See the top of Figure 2). The use in web search engines of the anchor text of hyperlinks on web pages as a representation of that web page, which receives the link, exploits the same idea (Brin and Page, 1998).

Aside from the author, another major cognitive agent who typically produces representations of academic documents is the *human indexers.* Their cognitive structures are represented by the index terms added to the original documents. Essentially they are the result of an intellectual interpretation of an author's text or images, often guided by predefined rules. Indexers typically select class codes and descriptors. These are taken from controlled vocabularies, e.g., a thesaurus (for which other human beings are responsible). Again, a thesaurus is an interpretation of the vocabulary and semantic relations of concepts within the domain covered by the thesaurus.

Consequently, the thesaurus constructor restricts the interpretation of the indexer and has an indirect influence on the representations made by the indexer. Apart from the indexer representations shown in Figure 2, uncontrolled terms and phrases (often called identifiers) may also be assigned. The identifiers represent the indexer's cognitive structures more directly, because the indexer chooses them freely. The thesaurus, the left-hand side, Figure 2, may in addition serve as a device for providing lead-in search keys (Bates, 1986), and as an automatic search key expansion tool.

A final group of representations that are indirectly influenced by a variety of cognitive agents are the *selectors* mentioned above. They are shown in the lower left corner of Figure 2, and contain structures that are *selective* and different from those of indexers, authors and users. Instead of aboutness, such features reflect *isness*—that a document actually exists. Most of the common bibliographic data or metadata thus belong to representations of isness by making available non-topical features connected to documents, depending on media, domain, and presentation style. They are the result of selection or assessment processes performed by various agents on documents and their authors over time (Ingwersen, 2002).

Examples of these include journal editors or a conference committee responsible for items that are selected for publication in a journal (or conference). The editor and not entirely the author determines the publication year or date. Such agents are affected by their social/academic context over time. The authors' affiliations also possess selective power, e.g., by hiring particular researchers.

5. POLYREPRESENTATION OF THE IT COMPONENT OF IR SYSTEMS

The IT components in the information space consist of representation techniques and indexing rules, matching algorithms, database architectures, search languages and computational logics produced by systems designers and producers. That is, the IT components are the main interest of researchers in the mainstream system-driven IR tradition. A given *retrieval engine* is an implementation of a particular combination of these. From a cognitive viewpoint a retrieval engine is the embodiment of the ideas of its creators. Thus cognitive overlaps between *different* retrieval engines can be studied. A large volume of research within the system-driven tradition clearly shows that the various best match IR techniques retrieve different but overlapping results, and the more alike the retrieval algorithms, the larger the overlap.

Among others Rajashekar and Croft (1995) investigated combining results from multiple algorithmic and operational index representations and retrieval strategies (engines) by application of the plausible inference network model of IR. They found indications of performance improvements in some combinations. The ranking produced by one best match IR system is ultimately a cognitive representation of the knowledge structures of its designers, and hence of interest to the principle of polyrepresentation: The simultaneous application of *several* different engines is consequently assumed to provide overlaps of objects of superior value to searchers than each single engine alone. The idea of data fusion in IR, as explored, e.g., by Belkin et al. (1995), thus corresponds to aspects of polyrepresentation.

When the principle of polyrepresentation is applied for IR purposes, the various representations of the documents are matched with the representations of the user's information need by means of a retrieval engine. Figure 1 above illustrates the enriched platform with a range of possibilities for this match created by the principle of polyrepresentation. On the left hand side of the figure the objects containing information are shown as semantic entities (denoted by $S_1 \ldots S_n$), which may be whole documents or parts of documents[1]. Each of these may give rise to a number of representations with various cognitive and functional origins as discussed above ($r_1 \ldots r_n$) that may be matched with the functional representations of the user's information need, either the work task description (w), problems statements ($p_1 - p_n$) or requests ($q_1 - q_n$).

The w, ps and qs are in a form suitable for manipulation by the retrieval engine used, and there may be several versions of each as the cognitive space of the user changes over time. Note that Figure 1 demonstrates this for *one* retrieval engine only. As mentioned above, different engines, e.g., one based on the vector space model (Salton, 1989), and one on the inference network model (Turtle and Croft, 1990), will retrieve and rank different documents, and the cognitive overlap between these may also be worth examination.

6. PROBLEMS AND POTENTIALS

Compared with the mainstream IR research tradition and the research carried out in the information seeking community, the principle of polyrepresentation offers a much broader approach than either two. The mainstream system-oriented IR research tradition focuses on document and request representation and matching algorithms, but not on the actual users of the system (Baeza-Yates and Ribeiro-Neto, 1999). Similarly, the user-oriented information seeking community focuses on the user's situation and seeking behaviour, but rarely on the IR systems involved. The principle of polyrepresentation stresses the importance of all agents and the interplay between them as a condition for achieving successful and optimal IR.

The potential of the principle is therefore to serve as a common theoretical framework for research that integrates the information seeking perspective on the users with the mainstream IR focus on designing and testing better IR systems. In addition, the validity of the assumptions expressed in the hypothesis above in Section 2 is clearly demonstrated by available research from several sub-disciplines in information science.

---

[1] The left hand side of the figure is inspired by the inference network model (Turtle and Croft, 1990; Turtle, 1991)

See Ingwersen (1996, 2002) for a variety of examples from IR, information seeking, bibliometrics and scientometrics. The principle is not only a theoretical contribution, it also points to areas in which it can be applied and to research methodologies to be employed. This section examines the potentials and problems of putting the principle into practice, and points to challenges to be faced in the near future.

7. STUDIES OF THE POLYREPRESENTATIVE NATURE OF IR

Relatively few studies have been based explicitly on the principle of polyrepresentation. Also large-scale interactive IR experiments with polyrepresentation implemented both on the system and user side are still lacking. One important reason for this may simply be that the principle, because of its holistic and comprehensive nature, requires a very large research effort to implement in its entirety. This is in part due to the fact that so far there has been a lack of test environments that are suitable for working with all the elements in the principle. Methodologically the principle of polyrepresentation is complicated to work with because at least two to three variables and their interplay are studied at the same time, e.g. {retrieval engine, user, user's context} or {retrieval engine, interface, user} (Ingwersen and Järvelin, 2004).

For instance, the main tracks in Text Retrieval Conferences (TREC) are based on the laboratory model, where users are not involved interactively (see, e.g., Harman, 1993). The TREC test corpora are large, but consist mainly of news articles with very limited possibilities for polyrepresentation of the documents. The interactive track at TREC has demonstrated, however, that users can form part of IR experiments in controlled settings (Hersh and Over, 2001). Although polyrepresentation of the documents is hard to achieve with the current document collections in TREC, there is no reason why it should be impossible to experiment with polyrepresentation on the user side. Information seeking research, on the other hand, is based on studying the user, and involves to an extent, e.g., user interaction with operational online systems. The document surrogates in these systems offer some possibilities for polyrepresentation of the documents. Research from this perspective would therefore at first glance seem better suited for experiments with the principle of polyrepresentation.

The main difficulty here is for end users to juggle all the possible representations and overlaps manually. Selecting which representations to combine and carrying out the necessary and complex Boolean combinations may lead to cognitive overload—or simply prohibit searching. Automated interface functionalities, such as the request model builder (RMB) indicated in Figure 1 above, would thus be important for an operational system based on the principle of polyrepresentation. The RMB would question the user to extract several versions of the information need and the underlying work task, and assist in the matching of these with representations of the documents and prioritising them over the search session. These tools have, however, yet to be developed.

In the initial proposals of the principle of polyrepresentation (Ingwersen, 1994, 1996), the RMB was inspired by elaborate user modelling approaches such as those put forward in the MONSTART and MEDIATOR models (Belkin et al., 1987; Ingwersen, 1992). Very complex RMBs can be constructed from such models, but we propose to begin more modestly with simpler versions. The investigation of principles for how a

RMB should function is central for future research on the principle of polyrepresentation. One initial approach to test would be to attempt to acquire different descriptions of the 'what', the 'why', and the perceived work task from users over search sessions.

A simple interface could interview users, e.g., first asking for what is needed (like the requests submitted to the present search engines), then enquiring why the information is needed (i.e., a description of the current problem to be solved), and finally asking for a description of the broader work task from which the current problem has arisen. These descriptions could then be matched against document representations according to general models of information seeking behaviour. As indicated in Figure 1 above, the description of the perceived work task/interest is likely to remain stable (at least for the duration of a search session), whereas the current problem might change over the session, giving rise to a number of different requests.

These kinds of assumptions could be part of the matching model. Even if no interactive interface is available to match the extracted descriptions in a polyrepresentative manner, such descriptions could be used to *simulate* an interactive system. Initial simulation studies could test which methods would be appropriate for matching different representations of the information need with document representations. This is similar to the simulation approach by White et al. (2004) discussed below. An alternative source for such descriptions might be existing test collections:

From 2004 the IR topics of the INEX test collection[2] includes descriptions that are similar to the ones discussed above. The INEX 2004 topics are based on the idea of simulated work tasks as put forward by Borlund (2000, 2003): In order to be used in the interactive track of INEX, it was required that all INEX 2004 topics not only detail *what* is being sought for, but also *why* this is wanted, and in what *context* the information need has arisen. Simulated polyrepresentation studies can be carried out with static topics like these, although much of the realism would be lost due to the missing interaction. The results could inform us which features are important to include in an interactive RMB, and result in better prototypes to be used with real users.

More advanced RMBs could include the approaches to incorporating captured user behaviour into adaptive relevance feedback algorithms studied by Ruthven, Lalmas, and van Rijsbergen (2003) and White et al. (2004). Ruthven, Lalmas, and van Rijsbergen (2003) used a novel approach to *explicit relevance feedback*: information on *how relevant* the user finds a retrieved document is combined with *how old* the assessments are, giving priority to the most recent assessments. The explicit assessments are used to select terms for query expansion and adapts to the user's behaviour over the search session. White et al. (2004) studied how *implicit feedback* detected from browsing behaviour can be used as input to relevance feedback. The parts of documents the users choose to display and the order of these are part of the implicit feedback. They are used to adapt the query to the user's information need. Both studies demonstrate how better representations of the user's current information need can be obtained through interaction, and that a number of representations can be extracted over a session. These

---

[2] Every year from 2002 the Initiative for the Evaluation of XML Retrieval (INEX) constructs an IR test collection from 12,107 full text articles from the IEEE Computer Society's 20 journals. For more information see Gövert and Kazai (2003) and the INEX web site http://inex.is.informatik.uni-duisburg.de/2005/.

representations can be combined with the three different types of descriptions discussed above, in particular with the perceived work task description. Alternatively, the latest version can be used directly as an up-to-date representation of the request and/or the current problem to be solved.

8. THE MATCHING PROBLEMS IN CREATING COGNITIVE OVERLAPS

Even with a working RMB prototype that could interview the user and extract different representations of her cognitive space, the question of how to match representations and generate cognitive overlaps may not be straightforward. The hypothesis above in Section 2 offers some advice, but only of a general nature (i.e., to match the ones that are most different in cognitive or functional type—and probably also temporally different). We regard this issue as one to be solved largely by empirical research. Although the principle of polyrepresentation has not resulted in a large body of empirical research that deal with all its elements, a few recent studies have worked with some of them. Below, four such studies are examined. All four have implemented some form of polyrepresentation of the documents, but none have attempted to represent the user's cognitive space in several ways—only a single, static version of the information need was used. Two of the studies were carried out in operational online databases (Christoffersen, 2004; Larsen, 2002), while the other two used test collections (Larsen, 2004; Skov et al., 2004).

Larsen (2002) proposes a new strategy for searching via citations, the so-called 'boomerang effect'. The strategy was tested in a small experiment carried out in the online version of the Science Citation Index (SCI). One test person provided three information needs based on his current research interests as a researcher at a hospital. The retrieved documents were subsequently assessed for their usefulness in relation to the needs. As the SCI does not contain the full text of the documents the number of representations used was limited: titles and abstracts by the author, and Keywords Plus (automatically assigned identifiers) and the network of references and citations by other cognitive agents.

The experiment was Boolean because of the online setting, and only static versions of the information needs were used. Larsen used the principle of polyrepresentation as inspiration for selecting representations, for the automatic identification of seed documents for the citation search, and for refining the results of the citation search strategy. This is an extension of Pao's (1994) investigations.

In brief, the strategy extracted the references from document sets retrieved from several different representations, and identified seed documents from the cognitive overlap between these. The seed documents were then used as starting point for a forward chaining through the citation network to identify documents that refer to the seed documents. The results of the boomerang effect consisted of a number of retrieved document sets ordered in a polyrepresentative overlap structure. They showed that the overlaps generated by many representations consistently had higher precision.

Although it was possible to implement the strategy for the purpose of the experiment, Larsen notes that with more than three or four representations the number of overlaps and the effort required to handle them increase dramatically. He also experienced problems with the initial query formulations, which had to be expanded in order to fit the individual representations. The expansion was also necessary in order to ensure that the sets were sufficiently large to produce an overlap in the first place. Finally, the output was a semi-ranked list of document sets with no internal ranking. This can present problems to users if a set is large, and this kind of output makes it very hard to compare the strategy to other IR approaches, especially best match systems.

Christoffersen (2004) used the online versions of Medline, Embase and SCI to test the proportions of relevant documents in the overlaps between the three databases. The representations had strong cognitive differences: Title/abstract words (from authors extracted from Embase) vs. MeSH terms (from indexers extracted from Medline) vs. searching by citations (from citing authors in SCI). The relevance assessments were by subject experts, and the results showed that the degree of overlap (i.e., the number of sets a document appeared in) correlated strongly with the percentage of relevant items in a set. Again the study was Boolean because of the online setting, but as only three representations were used no serious problems were experienced in handling the overlaps. The intersections involved did, however, reduce the number of documents in the overlaps to less than 14% of the total number of documents retrieved. This demonstrates polyrepresentation as a strong precision tool.

Both studies were on a small scale and used only a few representations, and the promising results of both are therefore remarkable. This may be interpreted as a consequence of the strong cognitive differences between the representations used. An equally important factor is the size of the database involved: in both cases the systems were operational and contained several million records each. The principle of polyrepresentation could therefore reduce the uncertainty associated with each individual representation, and still create overlaps that were not empty.

Skov et al. (2004) set out to test elements of the principle of polyrepresentation in a best match setting. A small test collection containing 1239 Medline records augmented with references and citations, but without the full text of the documents, was used (Shaw, Wood, and Tibbo, 1991). Despite its small size the collection offered several cognitive and functional representations: words from titles and abstracts (from the author), Minor and Major MeSH headings (by indexers) as well as references and in-going citations (by citing authors).

Skov et al. (2004) selected four of the representations and generated all possible cognitive overlaps between them to study the performance of each one in their experiment. Two types of queries were tested in a best match system: natural language queries and highly structured queries. Both types used Boolean operators to identify overlaps. The highly structured queries also contained indications of query facets and phrases, and had synonyms added intellectually from MeSH. These additions were the results of Skov et al.'s efforts to, on the one hand, improve the quality of the document sets, and on the other, ensure that the overlaps were non-empty. Results showed that overlaps generated by several representations had higher precision than those generated from few representations for both query types.

Marked differences were also found between representations; in particular the results indicate an increase in precision when documents identified by a citation search strategy formed part of an overlap, stressing the importance of using representations that have strong interpretative differences. In all cases the highly structured queries achieved higher precision than the natural language queries, which is explained as a consequence of generating overlaps in a best match system: because the natural language queries only require one search term from the query to be present, the retrieved sets of documents, and thus the overlaps, may contain documents with very little relation to the information need. In contrast to this, the highly structured queries ensure that all facets of the queries are present in each cognitive overlap. Skov et al.'s (2004) results indicate that the quality of the initial sets from which overlaps are created can be improved and better results achieved with the principle of polyrepresentation, but only after extensive work on refining and expanding the queries. This can also ensure that recall does not suffer too much.

Larsen (2004) tested a best match version of the boomerang effect using the INEX computer science test collection. Because of the complex full text XML structure a number of functional representations could be extracted from the documents: title, abstract, author keywords, cited titles (from the reference list) as well as figure and table captions and the introduction and conclusion sections. In addition, the documents were represented by descriptors from the INSPEC thesaurus and uncontrolled identifiers assigned by INSPEC indexers.

The boomerang effect was tested against two baselines: a bag-of-words index where all the representations were mixed into one, and a polyrepresentation baseline, which gave higher weights to documents retrieved from several representations and required that documents be retrieved from at least two representations. The same unstructured queries were used in all three runs. Results showed that the bag-of-words baseline outperformed the other two, and that the polyrepresentation baseline performed slightly better than the citation search strategy in the boomerang effect.

Strict Boolean overlaps were not enforced in any of the strategies; Larsen had instead chosen to rely on thresholds to limit the size of the sets. It should be noted that the best performance of the latter two strategies were obtained at relatively low thresholds, i.e., when the sets retrieved from each representation contained few documents. This may be explained similarly to Skov et al's results: as the best match system only requires at least one of the query terms to be present in the retrieved documents, only the documents at the top of the rank have a sufficiently strong relation to the information need.

9. COGNITIVE OVERLAPS IN BEST MATCH IR SYSTEMS

The approaches in Skov et al. (2004) and Larsen (2004) both produced ranked output, which has the advantage that it can be compared to standard best match IR methods. Best match methods are not, however, unproblematic when trying to generate strong cognitive overlaps. Best match systems will most often place the documents that contain all the query keys at the top of the ranked retrieval output, but will also include any document that contains just one of the query keys at lower positions of the rank. In addition, if a query key occurs very rarely in the database, but very frequently in a particular document, this document will be placed in the top of the rank because of the *tf*idf* weighting scheme, regardless of whether it contains any of the other query keys. The combination of partial match and ranked output is one of the main advantages of best match systems over exact match systems (Belkin and Croft, 1987).

However, in relation to the creation of overlaps in the principle of polyrepresentation there is a risk that the quality of the sets that the cognitive retrieval overlaps are based on as a whole is too low. For instance, with two search keys there is the risk that only the first of them is retrieved by some of the lower ranking documents from one representation, and the second in the lower ranks from another representation. Therefore proper polyrepresentation in the true sense of the concept cannot be achieved.

The results by Skov et al. (2004) and Larsen (2004) demonstrate that unstructured applications of the principle of polyrepresentation are not likely to result in performance improvements—rather a decrease in performance can be expected, at least when simplistic fusion strategies such as those in Larsen (2004) are used. Thus, the implementation of the principle of polyrepresentation in best match systems is not straightforward. It seems that some structure like that offered by Boolean logic is needed to ensure the quality of the cognitive overlaps as seen in Skov et al. (2004). The intention behind the principle of polyrepresentation is, however, that it should be applicable in both exact and best match settings. This calls for further investigations in which best match techniques can be fruitfully adapted to work with the principle.

One candidate would be the inference network model proposed by (Turtle, 1991; Turtle and Croft, 1990). The basic idea in the model is to view the retrieval process "... as an evidential reasoning process in which multiple sources of evidence about document and query content are combined to estimate the probability that a given document matches a query" (Turtle and Croft, 1992, p. 280). One advantage of this model is that it can represent many IR approaches and combine them in a single framework. Such a framework will allow experiments that examine the effect of generating cognitive overlaps with a variety of matching methods, both exact and best match approaches and their combinations.

The aim will be to find matching methods that will allow ranked output and, at the same time, ensure the quality of the cognitive overlaps. Another line of research to follow is investigations similar to Skov et al. (2004), where all possible overlaps between a range of different representations are examined to identify the most promising combinations. This also includes investigations of which representations do *not* intersect. For instance, Skov et al. (2004) decided to fuse the titles and abstracts into one representation because they tended to retrieve the same documents in their experiment. This becomes even more important when richer document corpora are used, e.g., the INEX corpus where a large range of functional and cognitive representations can be generated from the documents.

As shown by Skov et al. (2004) it may also be important to attempt to improve the quality of the document sets before the intersection to generate a cognitive overlap. Therefore different query expansion techniques are an important and necessary issue in research on polyrepresentation. It is important to note that the different representations may require different kinds of expansion in order to be successful.

10. THE POLYREPRESENTATION CONTINUUM

As shown by the studies discussed above the matching of representations with different cognitive and functional origins of both document and the user's information need is a very essential research issue for the principle of polyrepresentation. The main challenge is to identify methods of achieving a flexible match of representations while still retaining some of the power and quality of the Boolean logics. The studies analysed above show that this is not straightforward, and that much more work needs to be done

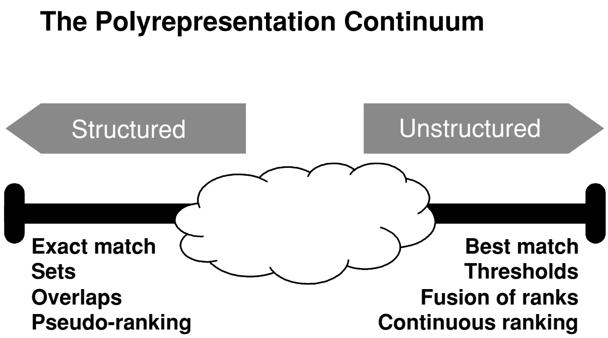

Figure 3: The polyrepresentation continuum (Larsen, 2004)

on the matching of representations before the potentials of the principle can be fully realised. Rather than using either exact match or best match approaches it is our belief that a combination of methods is needed.

Therefore we propose the idea of a *polyrepresentation continuum* as illustrated in Figure 3 below. The continuum is useful as a model for discussing how structured a given implementation of the principle of polyrepresentation is, and may guide the direction of further work on the principle.

At the *structured* pole of the continuum the implementations are based on exact match principles, leading to sets of retrieved documents for each representation from which overlaps can be formed and a pseudo-ranking be constructed. At the *unstructured* pole of the continuum the implementations are based on best match principles leading to a rank of the retrieved documents as input for polyrepresentation. Rather than generating overlaps between sets, the implementations at the unstructured end of the polyrepresentation continuum will fuse the ranks to produce a final ranked output, perhaps aided by thresholds to provide the necessary quality. Between the two poles there is a continuum of polyrepresentative constructs going from highly structured implementations to highly unstructured implementations.

The implementations in Larsen (2002) and Christoffersen (2004) are placed at the structured pole of the continuum. The polyrepresentation baseline in Larsen (2004) is placed at the unstructured end. The cloud in the middle of the continuum illustrates the current status where we do not know how to match the representations in a flexible and effective manner, and identifies the main challenge for future research on the principle of polyrepresentation: to identify flexible matching methods that can generate high quality cognitive overlaps from a variety of representations.

Skov et al. (2004) represents a constructive attempt to move from an exact match approach towards the unstructured pole of the continuum. Their highly structured queries that are run in a best match system can be seen as a second dimension of the continuum. Further moves toward the unstructured pole may include structured query types as investigated in best match systems by Kekäläinen and Järvelin (1998), where facets of the information needs are identified and expanded using a thesaurus. Query expansion and query adaptation to individual representations seem important and might lead to more formal IR models which incorporate differentiated normalisation and weighting for different representations. It may for instance be that the characteristics of certain representations are underrepresented by the standard *tf*idf* weighting scheme. Instead of calculating *idf* values in relation to the whole document, they might be calculated in relation to each representation in order to capture their individual characteristics more successfully.

In total we may observe at least three dimensions of the polyrepresentation continuum:

1. the degree of structure in the polyrepresentation matching modus, Figure 3;
2. the degree and nature of structure within queries by application of facets; and
3. the modus of expansion or modification of queries, like expansion of facets.

How influential the latter two dimensions are on the entire range of matching principles is yet to be investigated in a detailed manner.

11. CONCLUSION

The principle of polyrepresentation is a coherent and comprehensive cognitive framework that can be applied simultaneously to the cognitive space of the user and the information space of IR systems. The principle has the potential to guide the design of interactive IR systems that take full advantage of the available document representations *and* user's context to improve retrieval performance.

However, before this can be achieved a number of issues need further investigation. Among these are simulation studies that test which methods would be appropriate for matching different representations of the user's cognitive space with document representations. Such simulations could apply simulated work task situations; or they could explore exhaustively the possibilities of a number of controlled variables and thus simulate all achievable combinations. Investigations involving test persons and experimental laboratory tests (simulations) must take into account the dependency of domains, media and representation styles.

Studies of the principles for how a Request Model Builder should function would be fruitful, and how to match representations to generate strong cognitive overlaps, especially in best match settings. The latter issue is illustrated by the polyrepresentation continuum. It points to the investigation of flexible and powerful hybrid matching of representations as a challenge and opportunity for future research along these lines.

REFERENCES

Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Harlow: Addison Wesley. xvi, p. 513 ISBN: 0-201-39829-X.

Bates, M. J. (1986). Subject Access in Online Catalogs: A Design Model. *Journal of the American Society for Information Science*, 37(6), 357–376.

Belkin, N. J. (1980). Anomalous States of Knowledge as Basis for Information retrieval. *The Canadian Journal of Information Science*, 5, 133–143.

Belkin, N. J. (1984). Cognitive Models and Information Transfer. *Social Science Information Studies*, 4, 111–129.

Belkin, N. J. Kantor, P. Fox, E. A., and Shaw, J. A. (1995). Combining the Evidence of Multiple Query Representations for Information Retrieval. *Information Processing and Management*, 31(3), 431–448.

Belkin, N. J. Oddy, R., and Brooks, H. (1982). ASK for Information Retrieval: Part 1: Background and Theory. *Journal of Documentation*, 38, 61–71.

Belkin, N. J. Borgman, C. L. Brooks, H. M. Bylander, T. Crofts, W. B. Daniels, P. Deerwester, S. Fox, E. A. Ingwersen, P., and Rada, R. (1987). Distributed Expert-Based Information Systems: An Interdisciplinary Approach. *Information Processing and Management*, 23(5), 395–409.

Belkin, N. J. and Croft, W. B. (1987): Retrieval Techniques. In: M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST)*, 22, 1987 (pp. 107–145). Amsterdam: Elsevier Science.

Borlund, P. (2000): *Evaluation of Interactive Information Retrieval Systems*. Åbo: Åbo Akademi University Press. vi, p. 276 ISBN: 951-765-049-3. (PhD dissertation)

Borlund, P. (2003): The IIR Evaluation Model: A Framework for Evaluation of Interactive Information Retrieval Systems. *Information Research*, 8(3), paper no. 152. [http://informationr.net/ir/8-3/paper152.html, visited 6-8-2003]

Brin, S. and Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117. [http://www-db.stanford.edu/pub/papers/google.pdf (preprint), visited 16-5-2003]

Christoffersen, M. (2004). Identifying Core Documents with a Multiple Evidence Relevance Filter. *Scientometrics*, 61(3), 385–394.

De Mey, M. (1980). The Relevance of the Cognitive Paradigm for Information Science. In: O. Harbo and L. Kajberg (Eds.), *Theory and Application of Information Research: Proceedings of the Second International Research Forum on Information Science, 3–6 August 1977, Royal School of Librarianship, Copenhagen* (pp. 49–61). London: Mansell.

Dervin, B. and Nilan, M. (1986). Information Needs and Uses. In: M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST), 21, 1986*. New York: Knowledge Industry Publications, pp. 3–33.

Gövert, N. and Kazai, G. (2003). Overview of the Initiative for the Evaluation of XML Retrieval (INEX) 2002. In: N. Fuhr, N. Gövert, G. Kazai, and M. Lalmas (Eds.), *Proceedings of the First Workshop of the Initiative for the Evaluation of XML Retrieval (INEX): December 9–11, 2002, Schloss Dagsthul, International Conference and Research Centre for Computer Science*. [Sophia Antipolis]: [ERCIM], pp. 1–17. ([ERCIM Workshop Proceedings]) http://qmir.dcs.qmul.ac.uk/inex/Papers/final_overview_Goevert_etal.pdf, visited 27-7-2003]

Harman, D. (1993). Overview of the First TREC Conference. In: R. Korfhage, E. Rasmussen, and P. Willett (Eds.), In *SIGIR'93: Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 36–47). New York: ACM Press.

Hersh, W. and Over, P. (2001). Interactivity at the Text Retrieval Conference (TREC). *Information Processing & Management*, 37(3), 365–367. (Editorial: Introduction to Special Issue About the Interactive Track at TREC)

Ingwersen, P. (1982). Search Procedures in the Library Analysed From the Cognitive Point of View. *Journal of Documentation*, 38(3), 165–191.

Ingwersen, P. (1992): *Information Retrieval Interaction*. London: Taylor Graham. x, p. 246 ISBN: 0-947568-54-9. [http://www.db.dk/pi/iri/, visited 1-11-2003]

Ingwersen, P. (1994). Polyrepresentation of Information Needs and Semantic Entities: Elements of a Cognitive Theory for Information Retrieval Interaction. In: W. B. Croft and C. J. van Rijsbergen (Eds.), In *SIGIR '94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Organised by Dublin City University, 3–6 July 1994, Dublin, Ireland* (pp. 101–110). London: Springer-Verlag.

Ingwersen, P. (1996). Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation*, 52(1), 3–50.

Ingwersen, P. (2001). Cognitive Information Retrieval. In: M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST), 34, 1999–2000*, vol. 34 (pp. 3–51). Medford, NJ: Information Today.

Ingwersen, P. (2002). Cognitive Perspectives of Document Representation. In: H. Bruce, R. Fidel, P. Ingwersen and P. Vakkari (Eds.). *Emerging Frameworks and Methods: CoLIS4: Proceedings of the Fourth International Conference on Conceptions of Library and Information Science, Seattle, WA, USA, July 21–25, 2002* (pp. 285–300). Greenwood Village, Colo.: Libraries Unlimited.

Ingwersen, P. and Järvelin, K. (2004). Context in Information Retrieval. In: T. J. D. Bothma and A. Kaniki (Eds.), *Prolissa 2004. Proceedings of the 3rd Biennial DISSAnet Conference, Pretoria, 28–29 October 2004* (pp. 301–310). Pretoria: Infuse.

Kekäläinen, J. and Järvelin, K. (1998). The Impact of Query Structure and Query Expansion on Retrieval Performance. In: W. B. Croft, A. Moffat, C. J. van Rijsbergen, and J. Zobel (Eds.). In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR '98), Melbourne, Australia, August 24–28, 1998* (pp. 130–137). New York: ACM Press.

Kuhlthau, C. C. (1991): Inside the Search Process: Information Seeking From the User's Perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.

Kuhlthau, C. C. (1993). A Principle of Uncertainty for Information Seeking. *Journal of Documentation*, 49(4), 339–355.

Larsen, B. (2002). Exploiting Citation Overlaps for Information Retrieval: Generating a Boomerang Effect From the Network of Scientific Papers. *Scientometrics,* 54(2), 155–178.

Larsen, B. (2004). *References and Citations in Automatic Indexing and Retrieval Systems: Experiments with the Boomerang Effect*. Copenhagen: Royal School of Library and Information Science. XIII, p. 297 ISBN: 87-7415-275-0. (PhD dissertation) [http://www.db.dk/blar/dissertation, visited 11-2-2004]

Mackey, D. M. (1960). What Makes the Question? *The Listener*, 63(May 5), 789–790.

Pao, M. L. (1994). Relevance Odds of Retrieval Overlaps From Seven Search Fields. *Information Processing and Management*, 30(3), 305–314.

Rajashekar, T. B. and Croft, W. B. (1995). Combining Automatic and Manual Index Representations in Probabilistic Retrieval. *Journal of the American Society for Information Science*, 46(4), 272–283.

Ruthven, I., Lalmas, M., and Rijsbergen, K. V. (2003). Incorporating User Search Behavior into Relevance Feedback. *Journal of the American Society for Information Science and Technology*, 54(6), 529–549.

Salton, G. (1989): *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, Mass.: Addison-Wesley. xiii, p. 530 ISBN: 0-201-12227-8.

Shaw, W. M. Jr., Wood, J. B., and Tibbo, H. R. (1991). The Cystic Fibrosis Database: Content and Research Opportunities. *Library & Information Science Research*, 13, 347–366.

Skov, M. Pedersen, H. Larsen, B., and Ingwersen, P. (2004). Testing the Principle of Polyrepresentation. In: P. Ingwersen, C. J. van Rijsbergen, and N. Belkin (Eds.), *ACM SIGIR 2004 Workshop on "Information Retrieval in Context"*. Sheffield: [University of Sheffield], pp. 47–49. (Workshop proceedings) [http://ir.dcs.gla.ac.uk/context/, visited 7-2-2005]

Smeaton, A. F. (1992): Progress in the Application of Natural Language Processing to Information Retrieval Tasks. *Computer Journal*, 35(3), 268–278. [http://www3.oup.co.uk/computer_journal/hdb/Volume_35/Issue_03/, visited 17-2-2005]

Sparck Jones, K. (1990). *Retrieving Information or Answering Questions*. London: The British Library Board. (British Library Annual Research Lecture; 8)

Turtle, H. and Croft, W. B. (1990). Inference Networks for Document Retrieval. In: J.-L. Vidick, (Ed.). In *Proceedings of the 13th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: Brussels, Belgium, 5–7 September, 1990* (pp. 1–24). New York: The Association for Computing Machinery.

Turtle, H. R. and Croft, W. B. (1992). A Comparison of Text Retrieval Models. *The Computer Journal*, 35(3), 279–290. [http://www3.oup.co.uk/computer_journal/hdb/Volume_35/Issue_03/, visited 22-8-2003]

Turtle, H. R. (1991). *Inference Networks for Document Retrieval*. Massachusetts: University of Massachusetts. xiii, p. 198 (PhD dissertation) [http://ciir.cs.umass.edu/pubfiles/ir17.pdf, visited 4-6-2003]

White, R. W., Jose, J. M., Rijsbergen, C. J., and Ruthven, I. (2004). A Simulated Study of Implicit Feedback Models. In: S. McDonald and J. Tait (Eds.), *Advances in Information Retrieval, 26th European Conference on IR Research, ECIR 2004, Sunderland, UK, April 5–7, 2004, Proceedings*. Springer, pp. 311–326. (Lecture Notes in Computer Science; LNCS 2997) [http://www.cis.strath.ac.uk/~ir/papers/ecir_implicit.pdf (preprint)]

Wormell, I. (1981). SAP—A New Way to Produce Subject Descriptions of Books. *Journal of Information Science: Principles and Practice*, 3, 39–43.

CHAPTER 4

INTEGRATING APPROACHES TO RELEVANCE

IAN RUTHVEN
University of Strathclyde

1. INTRODUCTION

Relevance is the distinguishing feature of IR research. It is the intricacy of relevance, and its basis in human decision-making, which defines and shapes our research field. Relevance as a concept cuts across the spectrum of information seeking and IR research from investigations into information seeking behaviours to theoretical models of IR. Given their mutual dependence on relevance we might predict a strong relationship between information seeking and retrieval in how they regard and discuss the role of relevance within our research programmes.

However often, too often, information seeking and IR have been continued as independent research traditions: IR research ignoring the extensive, user-based frameworks developed by information seeking and information seeking underestimating the influence of IR systems and interfaces within the information seeking process. When these two disciplines come together we often find the strongest research, research that is motivated by an understanding of what cognitive processes require support during information seeking, and an understanding of how this support might be provided by an IR system.

The aim of this chapter is to investigate this common ground of research, in particular to examine the central notion of relevance that underpins both information seeking and IR research. It seeks to investigate how our understanding of relevance as a process of human decision making can, and might, influence our design of interactive IR systems. It does not cover every area of IR research, or each area in the same depth; rather we try to single out the areas where the nature of relevance, and its implications, is driving the research agenda.

We start by providing a brief introduction to how relevance has been treated so far in the literature and then consider the key areas where issues of relevance are of current concern. Specifically the chapter discusses the difficulties of making and interpreting relevance assessments, the role and meaning of differentiated relevance assessments, the specific role of time within information seeking, and the large, complex issue of relevance within evaluations of IR systems. In each area we try to establish where the two fields of IR and information seeking are establishing fruitful collaborations, where there is a gap for prospective collaboration and the possible difficulties in establishing mutual aims.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 61–80.

2. RELEVANCE

From its inception as a distinct research tradition, IR has engaged in a continuing discussion on the nature of relevance. The issues are large and significant: what does relevance mean as a concept, how does relevance manifest itself in online searching, and is relevance sufficiently robust to enjoy its predominant role in our understanding and measurement of IR systems? The latter is not a trivial point; several authors have argued that relevance is too weak and ephemeral to be used in understanding the human experience of searching.

However, as Saracevic (1996) notes, relevance has stuck as the core concept in IR research and this because relevance has many attractive features as a research tool: it is a concept that is intuitive to both searchers and investigators, it can be classified into different types and, within certain bounds, it is a concept that is measurable both quantitatively and qualitatively. Unlike other concepts that have been suggested as substitutes, e.g. information value or information quality, relevance does not require to be formalised within a specific information seeking situation to be of use, and unlike an evaluation metric such as searcher satisfaction, it is a measure that allows for a certain degree of repeatability across investigations.

Relevance emerges in many stages of the IR process: we use relevance to design retrieval algorithms, we ask users to provide assessments of relevance for relevance feedback and we exploit various measures of relevance to evaluate IR systems. This broad use of relevance means that relevance stands as the central force in our perception of the retrieval problem itself and concerns over the validity, reliability, and scope of relevance have been played out in the IR literature for decades.

Maron's (1964) early observation regarding the relationship between searchers and information indicates the nature of the relevance problem. "... information is not a *stuff* contained in books as marbles might be contained in a bag—even though we sometimes speak of it in that way. It is, rather a *relationship*. The impact of a given message on an individual is *relative* to what he already knows, and of course, the same message could convey different amounts of information to different receivers, depending on each one's internal model or map" (p. 9).

If, as Maron indicates, information is personal, the relevance of information is an even more personal notion. To help searchers find *relevant* information, our IR systems must deal with people in all their complexity: their tasks, knowledge, personality, motivation, etc. These variables are not simple and a simple notion of relevance does not suffice for the task of IR. This is why, the more we learn about the way people search and the way they assess information, the more complex the concept of relevance may appear. It is also why we have not agreed on a definition of relevance for IR.

This lack of consensus on a common definition of relevance may seem problematic; a science that cannot define its basic constructs would appear to be in trouble. However, relevance, rather than being a confounding characteristic, is an illuminating concept and one that helps us differentiate and appreciate the multifaceted nature of searching. It is not true as Mizzaro (1997) claims "relevance is not a well understood concept" (p. 810). Indeed, we understand it at many different levels and in many different ways, even if we do not properly understand yet what to do with all we have learned.

This chapter does not deal with definitions or reviews of relevance, as several good studies already exist in the literature. Borlund (2003b) for example, classifies and compares the main approaches to relevance, specifically for the evaluation of interactive retrieval systems, to demonstrate how many of the available definitions of relevance share a common background.

Mizzaro (1997) also tackles the issue of relevance from a broad perspective. In his article he surveys forty years of information science literature to overview how the concept of relevance has been tackled historically and to present the broad themes of investigation across the years. One of the main trends in this survey is the consistent call from researchers for understanding the nature of relevance as it impacts on human searching behaviour.

What is clear from these review articles is that relevance, in the wider context, is a subjective, multidimensional, dynamic and situational phenomenon. Relevance is subjective in that different searchers may have different opinions on the relevance of the same object. The assessment of relevance is affected by the assessor's background knowledge, motivation and perceptions of the assessment task itself, and even when given precise and strict criteria for assessing relevance, assessors' agreement on relevance can be low (Voorhees, 2000, 2001).

Relevance is multidimensional in that it is often not a simple binary assessment: one document[1] can be more relevant than another and an assessment of relevance can be based on more than one aspect of the document. Relevance is dynamic in that perceptions of relevance can change during a search session and across sessions. Indeed how could relevance not be dynamic? If one of the central purposes of a search system is to give searchers relevant information then the act of viewing relevant information must change the searcher's subsequent cognitive base for assessing relevance.

Both Borlund, Mizzaro, and also Saracevic (1975) in an older article, indicate that relevance can be measured as relationships between different entities, for example relevance can be measured against surrogates or whole objects, against a human cognitive need or against a concrete search request. Our measurements of relevance change as we change the method by which we measure it; that is, unless we have a solid background for investigation and measurement. It is not the case that we need *a* definition of relevance, rather we need to be clear about what definition(s) of relevance we are utilising in any investigation and how these fit into our understanding of relevance.

Historically, relevance has been tackled from two disciplines: information seeking (IS) and IR (IR). Relevance from the IS perspective has been concerned with the role relevance plays within the information seeking process, in particular understanding the types of relevance, levels of relevance or criteria for relevance. From the IR side, the concern has been how to use relevance for the design and measurement of systems. These two concerns have not always been aligned. For example, in IR evaluation the concern over relevance has often been more on the reliability of relevance measurements rather than the nature of the assessments. In this chapter, however, we hope to present a joint agenda for discussion and research on this important issue.

---

[1] As is common in the literature, the term document should be taken to subsume all searchable objects unless specifically stated otherwise.

3. RELEVANCE ASSESSMENTS AND INTERACTION

One of the core means by which we understand relevance is by the process of relevance assessment—the explicit or implicit act of distinguishing a document as being relevant. Obtaining relevance assessments whether to understand the judgement process, that is to understand how users are making relevance decisions, or as a means of gaining input to relevance feedback is of fundamental importance because the nature of relevance assessments reflects the nature of relevance: subjective, multidimensional, contextually situated and dynamic.

Relevance assessments are complementary in nature to the process of requesting information. When forming a search request we are asked to describe what information we want retrieved, usually without knowledge of what could be retrieved. Relevance assessments, on the other hand, are retrospective assessments on what has been retrieved. Analysing the reasons people give when assessing relevance has provided the basis for many detailed accounts of *criteria* for relevance. We know from these studies of relevance criteria, e.g., Barry and Schamber (1998) or Tombros, Ruthven, and Jose (2005) that users are not employing simple mechanisms when deciding on relevance. Rather they are making a series of multifaceted, task-dependent, interrelated decisions on different aspects of the objects retrieved, the information given by the objects and the user's own information state. A repeated feature of these studies is that there is a large, consistent set of criteria that are used by different user groups in different search situations for assessing relevance.

Investigating these criteria, understanding the human decision-making process when accessing and using information, has always been a central topic for information seeking research but the implications have not always impacted as greatly on IR research as might be expected. IR systems still operate largely on topical relations between queries and information objects with little incorporation of a more diverse set of relations into the retrieval process. The recent impetus on incorporating meta-data, especially in web applications and in digital libraries, has been seen as a step in this direction although the work has largely tended to concentrate on objective data such as authorship, date, and content information, largely bypassing the more difficult subjective issues such as information quality, reliability, and novelty—decisions on which we know users are making when assessing information.

How we incorporate the ability to handle such diverse criteria for relevance into IR systems needs some thought. It is not enough to present more complex querying mechanisms because, although users can describe their criteria for relevance *after* seeing relevant material and with respect to the material actually retrieved, they cannot always verbalise or express these criteria before seeing retrieved material. Indeed, the gap between what searchers *expect* to be retrieved when forming a search request and how they retrospectively judge relevance is worth investigating to assess the degree to which reported relevance criteria are influenced by the methodology used in the investigation itself.

Although users can be sophisticated about their decisions regarding relevance, the interfaces developed for information access systems, in particular, still supply limited support for the range of assessments users make about retrieved material. Where IR interfaces *have* considered relevance assessments is as the mechanism to provide input to relevance feedback algorithms, the process by which a system uses identified relevant documents as the basis for modifying the user's query. Diane Kelly covers relevance feedback in detail in a separate chapter so we shall not duplicate this effort here, except to discuss the effect of interaction on the process of making an explicit relevance assessment.

3.1. Explicit Relevance Assessments

If we are to obtain relevance assessments from a user then we must consider how the assessment gathering mechanism should be included in the interface. Classical approaches to relevance feedback have required users to indicate the document in some way, such as ticking a checkbox. There are advantages to this lightweight approach: it is a conceptually, and physically, simple action to perform, and does not force a user to define what they mean by relevant. However, unless the system *forces* the user to make such assessments, e.g. as in Aalbersberg (1992), the act of making a relevance assessment is usually tangential to the user's main goal of finding information. The design of most search interfaces allows the user to assess a document but does not indicate what the system will do with the document, nor what difference will arise to the search results if the user does/does not assess a document as being relevant. That is, the interface permits the user to assess documents but does not support the decisions the user makes about relevance; neither does this interface design encourage users to make relevance assessments.

The design of typical relevance feedback interfaces is one in which the user collects relevant information and, at some point, asks the system to do something with it. This fits with the statistical generalisation approach to relevance taken by most relevance feedback algorithms; they generally work best with more examples of relevant information. This, however, would appear to be a system-driven design ('what evidence do relevance feedback algorithms need') rather than a user-centred design.

Laboratory studies of systems that offer relevance feedback demonstrate that users *can* employ relevance feedback successfully, especially in experimental settings where the concept has been explained, demonstrated and subjects have been encouraged to use the feature. However, it has not been demonstrated that relevance feedback is a technique that users actually want or will use operationally. The current evidence on operational trials of relevance feedback is patchy but generally indicates that the basic usability problems regarding relevance feedback have not been tackled in a meaningful way: users are still unsure about how to make feedback decisions, usually uncertain about the benefits of giving feedback, and often make poor decisions (Beaulieu and Jones, 1998).

3.2. Implicit Relevance Assessments

The major recent attempt to bypass the usability problems associated with relevance feedback is to develop systems that avoid the user having to provide explicit relevance assessments. These systems, *implicit* feedback systems, remove the difficulty of making explicit relevance assessments by making inferences from the users' interaction with information objects. Distinct actions such as printing, book-marking and inferred expressions of interest such as viewing time and scrolling behaviour are taken to represent interest in information provided by the objects. Implicit approaches have the major advantage for IR system designers that implicit systems can collect much more information from users than explicit feedback systems.

Implicit approaches can, potentially, also form the basis of a more holistic approach to modelling search behaviour by considering the temporal, linked series of actions a user performs when interacting. This is opposed to the explicit relevance feedback model where the input is a series of assessments with no indication of what happened in between assessments. At present, however, the state of the art in implicit modelling approaches still tends to use discrete pieces of behavioural information; the limitations of this discrete approach are demonstrated in the inconsistency of variables, such as viewing times, when measured against other variables such as user experience, and task knowledge (Kelly and Belkin, 2004).

3.3. Understanding Relevance Assessments

Implicit systems do not separate the act of making a relevance assessment from the general interaction: unlike in explicit feedback systems there is no distinct act that corresponds to a judgment of relevance at least for the purpose of communicating relevance to the system. Consequently, the user should be assumed to be communicating information that is indicative of their current interest; whether or not this information can be treated as relevant information is to be tested. Similarly with other attempts to use implicit feedback such as those systems that use perceptual stimuli, e.g. eye-movements (Salojärvi et al., 2003), we need to be sure that these stimuli are indeed reflecting relevance and can be treated in the same way as relevance. A user's intention when interacting with the system, including when making relevance decisions, is necessary for understanding what to do with the data detected by the system.

Similarly, it is worth reiterating the point that, implicitly or explicitly, relevance assessments are made on objects such as documents themselves, not on the representations of the objects being used for matching or the matching algorithms used for retrieval. We should be conservative about what inferences we can draw from simple assessments of relevance unless we learn how to use more of the factors that can affect these assessments.

For example, the newer environments in which online IR systems operate offer a valuable opportunity to investigate how criteria for relevance are affected by the context in which the system is used. In mobile information seeking, for example, where the user is not in a fixed location, the assessments a user makes may be influenced by the dynamic *physical* context of the user. The factors that affect relevance in newer situations such as these may be very similar to standard environments but the balance of what criteria are used or what types of relevance predominate may be very different. We do need more empirical work to assess how the context of interaction affects relevance assessments so we can make better, and more confident, inferences about what a user might intend by their interaction. We also need to understand which variables within investigative situations are having an effect on the nature of relevance assessments made by searchers. The difficulties of scoping out these research questions is shown next in a discussion on the deceptively simple issue of interpreting relevance assessments.

4. DIFFERENTIATED RELEVANCE ASSESSMENTS

Interfaces that allow users to make relevance assessments, either as input for relevance feedback or for investigative purposes, typically ask users to make a binary judgement on the relevance of a complete entity. Although this makes the use of relevance easier it has long been held to be a poor estimation of the decisions that users make and the issues involved in utilising differentiated relevance assessments have recently achieved a new momentum.

Differentiated relevance assessments can either be taken to mean

- Assessments that are made on a non-binary scale, *non-binary* assessments,
- Assessments that measure some group agreement on the relevance of an object, *consensus* assessments,
- Assessments that are made on the amount, or part, of a document that is relevant to the user's information need, *completeness* assessments.

Investigating these types of assessments reflect quite different avenues of research and, in the following sections, we shall look at the ways in which these types of assessment are being tackled.

4.1. Non-binary Assessments

Most authors on relevance acknowledge relevance to be non-binary in nature, i.e. a document is not simply relevant or non-relevant to an information need. Although the binary/non-binary distinction is common, particularly in IR evaluation, there are many arguments against such a simple use of relevance. A strong argument is that it is methodologically inadequate. Eisenberg and Hu (1987) indicate that simple relevant/non-relevant distinctions between items can distort standard evaluation criteria such as recall and precision because the *break-point*, the point at which users will make the distinction between relevant and non-relevant are not consistent. That is, the point at which a searcher will judge an item relevant rather than non-relevant is not a simple factor of the document content. Janes (1991) in a continuation study indicated that the familiarity with the topic, in particular, can affect where the break-point is drawn. Therefore, when using a binary scale we are not measuring relevance but relevance within a context framed by the investigation and searcher's knowledge.

How we allow users to quantify non-binary relevance, and how we interpret non-binary assessments, is not trivial nor are there mechanisms that are clear and unbiased. Mizzaro (1999) points to several means of eliciting non-binary relevance assessments. For example weighted assessments in which searchers are asked to assign a value to a document giving a relative relevance score, although it is not clear to what these scores are relative. Alternatively, we may employ scalar approaches in which documents are given a value on an ordinal scale, e.g., 1–7; an approach that subsumes the more general region-based approach of dividing assessments into groups: non-relevant, partially relevant, relevant.

Others have argued that subjects find it easier to order documents by relevance but, as Katter (1968) pointed out, ranking or preference based orderings may force searchers to make distinctions between the relative relevance of documents where the searcher sees no difference. We can also be more precise about the nature of relevance. Rather than simply presenting relevance degrees, Oard et al. (2004) describes a test collection that includes different reasons for relevance, e.g. object provides direct evidence, object is useful as basis for comparison, object provides pointer to relevant information so the assessor can differentiate reasons, rather than simple degrees of relevance.

Irrespective of how users make non-binary assessments, an interesting question is what do non-binary assessments actually represent? In part, this depends on what task the assessors are asked to perform and the situation in which the subjects are asked. There is, however, a growing distinction in the relevance literature between assessments that reflect uncertainty on the part of the assessor regarding relevance and a confident assessment that the document is only relevant to a degree.

For example, Spink, Greisdorf, and Bateman (1998) in a series of studies investigated the assessment of non-binary relevance across time. Two findings are of particular interest. Firstly, users who have less background knowledge make more non-binary relevance assessments and, secondly, non-binary relevance assessments were correlated with changes in relevance criteria: when users were progressing in their task they made more non-binary relevance assessments. The understanding put forward is that highly relevant documents may equate to documents that confirm the searcher's current state of knowledge whereas high numbers of non-binary relevance assessments relate to change and a new understanding of the task the user is trying to complete. Non-binary assessments in this case were not an assessment of the relevance of the information. Rather they reflected a feature of delayed decision-making: the documents assessed may later be assessed as more highly relevant when the user understands more about their task and how the documents fit into the task.

Choi and Rasmussen (2002) also demonstrated this aspect of delayed decision making in image retrieval; in their study, one reason given for assessments of non-relevance was user uncertainty as to whether an image would be ultimately be used. This uncertainty arose, in turn, because the nature of the task may change as the search progresses. They also noted, as the user's search progressed, that the user's relevance criteria changed, for example the availability and visual appeal of the images became more important whereas the importance of simple topicality decreased.

An assessment of the degree of relevance is not, therefore, just a simple decision on how relevant is a document but how relevant it is relevant to the user's current understanding of the role a document may play in completing a task. When making an assessment of high relevance the user is usually more able to make judgements on different criteria and different aspects of relevance such as pertinence, and utility (Choi and Rasmussen, 2002, Greisdorf, 2003) Uncertainty regarding these aspects is often a cause for assessments of lesser relevance.

These studies are methodologically different from studies where non-binary assessments are indications of the *degree* of relevance of documents either to each other or to the user's information problem. In these studies the assumption is that users can make definite assessments on the relative relevance of documents. Similarly, Sormunen (2002) and Vakkari and Sormunen (2004) directly assert the principle that users want highly relevant documents, documents that can objectively be assessed as being mostly composed of relevant information, and that we should base our evaluation on a system's ability to retrieve these documents.

Some attempts have been made to incorporate non-binary assessments into the retrieval process when taking non-binary to mean degree of relevance. Ruthven, Lalmas, and van Rijsbergen (2002) for example, presented one method of incorporating partial measurements of relevance assessments into standard relevance feedback algorithms by treating a partial relevance assessment as a part of a whole relevance assessment. In this case the documents with higher relevance scores were preferred as sources of query expansion terms. This approach could be challenged on many levels, e.g. several partly relevant documents may not be as useful as one highly relevant document, although empirical work does allow us to investigate the consequences of non-binary assessments within IR systems.

What we do with these two kinds of non-binary relevance would appear to be quite different: relative relevance assessments could be used to differentiate amongst sources of relevance evidence, whereas non-binary assessments caused by user uncertainty would appear to argue for dynamic interface support. Although it raises methodological problems, what, in fact, we may want to gather is not just an estimate of the degree of relevance of a document but also an estimate of the user's confidence in their assessments.

4.2. Consensus Assessments

In an early article Katter (1968) noted that a "recurring finding from studies involving relevance judgments is that the inter- and intra-judge reliability of relevance judgments is not very high" (p. 1). The concern about this lack of consistency is that if relevance assessments are unreliable, either across people or across search situations, then we cannot rely on any predictions or conclusions based on the assessments.

A possible solution to this problem is to develop methods of obtaining more stable assessments. For example, instead of using individual relevance assessments the process of assessment can be a collaborative process in which a group of judges decide on the relevance of an item. The result would be a *consensus* relevance decision, e.g. the item is considered to be relevant if all/most judges agree that it is relevant to the task or topic, or a graded assessment, e.g. each document has a relevance score indicating the proportion of judges who assess it to be relevant. This approach fits with Lesk and Salton's (1969) claim that highly relevant documents are the ones that most people would agree are relevant.

Campbell (2000) utilised the second approach, graded consensus assessments, to create an image test collection in which four independent assessors assessed each image to give four levels of relevance. The same test collection could then be used to test different scenarios: only using images on which all judges agreed on relevance, most judges agreed, or any item on which one judge claimed relevance. In a user test, Campbell's collection gave quite different performance figures for different agreement scenarios even though the scenarios appeared to retain the same relative ranking of alternative systems.

Although group assessments have been less studied historically, except when studying the consistency of assessment, they have gained a new impetus in the area of collaborative filtering. Collaborative filtering, often presented as a special case of implicit feedback, makes recommendations of items based on group consensus of rated items. Here the assessments are generally based on preference or taste but these can be seen as specific criteria for relevance, e.g. affectiveness or entertainment value (Barry and Schamber, 1998). Often collaborative systems allow explicit ratings, reflecting the fact that user's direct evidence is often better than implicitly gathered evidence, and this is one case where non-binary assessments are made as a matter of course.

Handling group assessments raises interesting questions for the combination of assessment; for example, should some judges be preferred over others, should we try to normalise judgements across judges, should judgements be considered only on an individual object or over a range of assessments made by an individual? Herlocker et al. (2004) in an excellent survey of evaluating collaborative filtering systems shows how many of the features associated with relevance, e.g. satisfaction, novelty, quality, are important in collaborative filtering and many of the issues of relevance from IR are important. In particular, the representative nature of the test collections used in the evaluations—that they are realistic data sets—is of critical importance; this is perhaps especially true for collaborative filtering approaches where the notion of a *user* is stronger than with an IR test collection which has a weak notion of individual users.

4.3. Completeness Assessments

Information objects need not be completely relevant; documents for example can contain relevant and non-relevant parts. A differentiated relevance assessment can therefore reflect a judgement of incompleteness, a measure of how much of the object is relevant. This fits with the partiality of finding information; the information we want may be tangential to the information objects with which we are interacting and no document may be completely about the topics for which we are searching.

INEX, the initiative on creating structured document test collections and retrieval system operates on both principles of non-binary assessment and incompleteness (Kazai, Lalmas, and de Vries, 2004). In INEX, document sections are assessed both on how much relevant material they contain (*relevance*) and how much of the section is devoted to relevant material (*exhaustivity*). This two-dimensional classification incorporates two aspects of differentiated relevance assessments: good sections are those that discuss relevant material in more depth, measured by the relevance criterion, and discusses less non-relevant material as measured by the exhaustivity criterion, both criteria being measured on a four-point scale. The ideal section will be both relevant and exhaustive.

This multidimensional approach to relevance assessments brings up interesting questions such as should we prefer sections that are exhaustive but not highly relevant or sections that are relevant but not highly exhaustive? The intuition would surely be that this is a user decision but there may not be a clear, consistent answer and straightforward interpretations of relevance metrics such as recall and precision are insufficient to answer these questions.

This becomes more complicated when dealing with more complex retrieval situations. When dealing with document sections, as in INEX, the question arises: if a relevant document section has been retrieved then do we count sections that contain or are contained within this section as also being relevant? This issue of inherited relevance is not restricted to textual retrieval; any media type such as video retrieval that has some notion of structure is affected by the issue (de Vries, Kazai, and Lalmas, 2004). Our decisions on how to tackle these questions are based on our use of relevance: counting these sections as non-relevant might distort our evaluation of the retrieval algorithms but be counterintuitive to a user who might expect relevant information to be novel.

Kazai, Lalmas, and de Vries (2004) discuss some of the possible ways to model different evaluation metrics for assessing the performance of structured document retrieval systems under different search scenarios. What is significant about these modelling approaches is that user satisfaction, as well as relevance, is a predominant concern: what would a user want from such a system and how should this be incorporated within an evaluation?

5. SUMMARY

Allowing users of IR systems to make differentiated relevance assessments would seem a simple extension to the standard IR interface. However, as has been seen, we cannot make simple inferences based on these assessments without considering why the assessments may have arisen and what they indicate about the user's search. The concern regarding how the user indicates differentiated assessments at the interface may be the wrong concern. Allowing the user to differentiate the way in which documents are relevant may be a more useful approach.

For example, it may be the case, as Greisdorf (2003) indicates, that users can and do make *more* decisions about the relevance of documents they judge as being highly relevant, that relevance is a more multifaceted decision than non-relevance and that the system can gain useful information by being told which criteria are important. Support for these findings was given by McLaughlin and Sonnenwald (2002) also looked at the criteria for relevance, especially for partial relevance. They showed more criteria being used for assessments of highly relevant documents and more positive criteria (features of the document that were present). For partially relevant the criteria used were more mixed (positive and negative criteria).

These aspects of making relevance assessments cannot simply be subsumed under a general relevance scale. As in the previous section, simple methods of interaction may fit with the requirements of those who design retrieval algorithms but they do not allow us to properly utilise the rich evidence that searchers can supply as evidence to these algorithms.

6. RELEVANCE AND TIME

Relevance is not a fixed commodity and, as outlined in the previous section, a user's perception of the relevance of an object can change over time, as can the relevance decisions a user makes about an object. IR systems have responded poorly to this phenomenon with user adaptation either being absent altogether (as in most web searches) or being restricted to simple monitoring of the topics for which a user is searching.

One of the few attempts to explicitly recognise time within IR systems was made by Campbell (2000) with the notion of ostensive relevance. The ostensive relevance of an object is seen as how representative of relevance is a current information object. That is, as relevance criteria may change over the course of a search, objects marked relevant at the start of a search should be seen as less indicative of relevance than objects recently marked relevant. In Campbell's work this was used to develop a retrieval model that biased query modification in favour of those objects most recently marked relevant by the searcher.

Neither has IR responded to the linking of searches often seen in information seeking. Users often repeat searches, perhaps using the same topical query but requiring different information returned. Spink (1996) for example, reports on a study of when and how academics use IR systems over the course of a research project. The majority of academics searched at the beginning of project and many search again throughout the project. One reason for searching at later stages of projects is to check new updated references—rerunning same searches against new data—but many users modify their search terms over time, either as their information problems change, and they need different information, or they obtain information from new sources. Although the searches seem similar on the surface and the basic topics of the searches are broadly the same, the reasons for searching and the type of information being sought is different leading to different relevance assessments.

Vakkari (2000) also looked at the changes in search behaviour in a longitudinal study of writing a research proposal showing an increase in the use of search operators, increased specificity of terminology and more sophisticated use of search tactics at the end of a search when subjects were more informed and had a clearer idea of what they were looking for. These are particularly appealing results because we have a set of behaviour-based observables: measurable units that can potentially be incorporated into an adaptive search system.

These studies have typically involved periods where the user has run multiple related, but distinct, searches over a period of perhaps weeks with intervening periods where their information problems can develop and mature. However the same change in need and understanding can also be seen in relatively short web searches, e.g. Tombros, Ruthven, and Jose (2005). As noted before, the fact that we can establish that different criteria are being used in different searches and that these criteria change over time makes it a pity that few attempts have been made to utilize this important facet of human search behaviour within IR systems. Attempts certainly have been made to track user interests through user modeling techniques; see (Ingwersen, 2002) for a detailed discussion for example. However, the notion of trying to dynamically map user criteria for relevance within related searches has not been significantly explored within practical implementations.

Detecting the reasons for relevance was one of the main original aims of the work on logical modelling of IR systems, i.e. equipping systems with the ability to state multiple relations between objects and specify which relationships correlated with relevance. However, the fact that even this branch of modelling has not produced any real results in this area is disappointing because the work on criteria for relevance can essentially be seen as setting the goals for a reasoning approach to modelling IR systems. Reasoning about what type of information may be required, based on what we already know about user's previous interactions and relevance decisions, would appear to be a rich area for IR systems research due to the many research questions it throws up: what units of time are effective for making decisions, how do we know if users want different information in this search, how much do users remember about their previous searches or how they previously measured relevance?

7. RELEVANCE AND EVALUATION

The study of relevance within evaluation is one area where information seeking research has typically provided strong empirical evidence on the nature of relevance and the factors that affect assessments, but where IR has been slow to exploit these findings. Relevance, and metrics based on relevance, has been at the heart of IR evaluation for decades and it is in this area where the limitations of a simple view of relevance have been most clear.

It is worth reiterating the general point that any evaluation approach tends to distort what it tries to evaluate. Evaluation as an activity highlights some aspects of the phenomenon being studied and ignores others. As Hersh and Turpin (2004) demonstrated, employing simple relevance metrics in user evaluations can give misleading results because simple metrics may ignore the factors that influence relevance decisions. The important point is that how we characterize, measure and use relevance in an evaluation, implicitly or explicitly carries a definition of success. This definition of what it means to succeed in an evaluation carries with it, in turn, the definition of what we see as the task of IR systems.

For IR systems evaluations, test collections have become the major instrument for evaluation. The TREC initiative set a standard for test collection construction that has been influential in creating many of the test collections in use today. The elements are: a set of information objects, a set of requests and a fixed set of relevance judgements made on all, or a sample, of the information objects. As the major tool for IR system evaluation, test collections have the major advantage of repeatability across time. They also have the advantage, difficult to replicate in user evaluations, of giving the ‘what-if’ experience: what if some component of the IR system were different, what would be the effect on the retrieval decisions made by this system and how might this affect our assessment of the system?

Voorhees (2000) demonstrated the stability, and hence reliability, of test collection evaluations under different sets of relevance assessments. Her investigation indicates that, although relevance assessors may disagree on the relevance of individual items, the ranking of IR systems based on a set of relevance judgements is not significantly affected by this disagreement. This is with the caveat that we need a sufficiently large number of topics to make this true. One of the main factors in the stability of test collections was that documents that were unanimously agreed to be relevant were ranked higher and so counted more in measuring the stability of the test collection. This would appear to fit with Lesk and Salton’s (1969) claim, noted earlier, that there will be more consensus regarding highly relevant documents and certainly this might be true if we concentrate on making topical relevance assessments as is the case in TREC.

Even in a controlled environment, however, subjective factors do play a role in the stability of test collections, e.g. in the Voorhees study she indicates that the assessor who created the topic, and possibly knew more about the topic, tended to judge more items relevant than neutral secondary assessors.

Test collections historically have attracted much criticism, partly due to their reliance on binary and topical notions of relevance, elements that are perhaps more easy to overcome than would appear from the lack of collections that incorporate differentiated assessments or non-topical assessments. A more significant criticism regarding test collections is that, although there are many test collections for different media types, generally test collections lack diversity in their construction and are lacking in what information an evaluator has to work with regarding the relevance decisions made. We have queries, assessments and documents but little else regarding the context of these artefacts such as the context of the creation of the queries, the order in which assessments were made, or the assessor's satisfaction with the documents assessed.

Neither do we have as a component in the test collection itself knowledge we know to be important in understanding the results of a search such as assessor experience, knowledge, or whether the search request was part of a global task. Although this information may be recorded at the time of creation of the collection, we cannot utilise it in interpreting the success or failure of a retrieval experiment. Multimedia collections are often richer in specifying the criteria used for assessment because the multivariate nature of the information objects, e.g. structural, visual, and content relations, are move obvious. However, even these collections often only record the assessments not the contexts within which assessments were made.

Harter (1996) specifically criticised the standard test collection model of evaluation because it ignored the variation in these contextual factors as to why relevance judgements may be made. His major argument is that variations in the factors involved in assessment are smoothed out by the aggregation involved in test collection methodologies meaning that we lose important results about success of individual queries and that we need to develop evaluation measures that are more sensitive to differences in why relevance assessments occurred. In particular we could argue that marginally relevant documents may serve useful purposes if we could establish why they were relevant to individual assessors.

Some attempts have been made to widen the scope of test collections. Reid for example, argued that *task* should be the basis for assessing relevance and constructing test collections (Reid, 2000). She suggested changing the focus of test collection construction from defining a relationship between a query and a document to the relationship between completing a task and a document, i.e. the assessment should be based on how useful the document was to completing a task. This means the assessment of relevance, for the test collection, should be made to the time at which the document became useful to the task to which the user is going to use information rather than the time at which the user is seeking information. Markkula et al. (2001) also looked at incorporating tasks into test collection construction for image retrieval and showed some limited success in improving the consistency of relevance assessments by use of task information: users, although not significantly, agreed more on the relevance of documents to a task than when using a simple search request.

However, even though test collections are essentially simulations of a potential user (what queries would a user submit based on the search topic, what documents would they be presented with and what documents would they assess as being relevant), we still cannot easily use these tools for assessing issues such as dynamism of relevance, or modelling the interactive nature of information seeking. Nor yet can we establish an answer to the question posed by Oddy (1981): is relevance in the laboratory the same as relevance in the real-world? In effect, does the use of relevance in a test collection relate in any way to how relevance might manifest itself when decided upon by a user in a specific situation?

Relevance itself is the basic criterion in most user evaluations of IR systems. In this case, we have the basic question of whether to use fixed relevance assessments, that is asking users to find the documents we have previously assessed as being relevant, or allow users to make personal relevance assessments, either on real information needs or given needs. The advantages of using previously defined relevance assessments are that we can identify for what types of relevant documents the system is performing poorly, that we can perform some error checking as was done in the interactive track of TREC, and we can define alternative evaluation measures, e.g. Sweeney and Crestani (2003) used *decision correctness*: the proportion of relevant and non-relevant documents correctly identified, a measure that is only possible when using ground truth test collections.

However there are many disadvantages, including the inherent variability of relevance assessments. Voorhees (2000) pointed to situations where the relevance assessments in test collections—the ground truth we use for evaluation—are less robust, namely situations where we have few queries, few relevance assessments and employ different criteria for relevance within the evaluation. Unfortunately these are the situations we encounter in user evaluations; experimental subjects run few queries, often have time to make few relevance assessments and will have different criteria for relevance than the original test collection assessors.

An alternative, of course, is to allow users to make their own relevance assessments. This means that we allow experimental subjects to develop their own criteria for relevance usually based on a search task. Borlund (2003a) and others have argued for this approach on the basis of its psychological realism and that it encourages situational, rather than simply topical, relevance assessments. Using situational relevance, or task relevance, as the basis for relevance assessments might make the evaluation more realistic but does not necessarily make it more reliable because the assessments are affected by the environment that is set in the study: how long users are given to search, what form of output they are set and also the modality of the output.

Tombros and Crestani (2000) showed less consistency in assessments when retrieved documents were spoken to experimental subjects than when the subjects were allowed to read documents themselves. Ruthven (2001) showed that the consistency of situational relevance assessments depends in part on the number of topically relevant documents in the collection, and Vakkari and Sormunen (2004) showed that the presence of few relevant documents encouraged users to employ more liberal relevance criteria than when there were many relevant documents available for retrieval.

In general, the experimental setting and its affect on relevance assessments is not properly understood. A particularly difficult issue is how to obtain relevance assessments in a neutral, objective manner. For example, it has long been established that one of the main factors in assessing the relevance of documents is the *order* in which documents are shown to the user. Several studies, e.g., Eisenberg and Barry (1988), Florance and Marchionini (1995), and Huang and Wang (2004) point to the fact that users assess the relevance of documents relative to the relevance of documents they have already seen, that assessments are relative to how much information is available, and viewing one relevant document can change the user's perception of the relevance of subsequently viewed documents.

It is not even clear whether we should treat all relevance assessments as a single set of assessments. Relevance feedback algorithms, with a few noticeable exceptions, treat relevant documents as a coherent set of items. However, users, especially if they cannot differentiate why documents are relevant at the interface, may not intend documents to be treated by the system in this fashion. Often the means by which we ask for relevance assessments at an interface reflects the system designer's view of relevance not a user's judgements. The issue of whether we treat a relevance assessment as a series of individual assessments or a group of related assessments is not trivial because relevant documents can interact when a human is assessing relevance; Tiamiyu and Ajiferuke (1988) for example, looked at the interactions of information between documents. They suggest three types of dependence can exist in retrieval situations

- *Independence*. Each document marked relevant should be treated as an independent relevance assessment, reflecting a decision made by the user on the information contained within the document.
- *Complementarity relationship*. The information contained within two documents sums to more than the sum of relevance ratings of each document together. An example would be the case where the relevance of one document only becomes apparent by reading another document.
- *Substitutability relationship*. The information in one document can substitute for the information in another document. This can cover the situations, for example, where we have duplicate or near duplicate documents in the collection being searched.

They show, theoretically, that the presence of different types of relationships can, although giving the same results for evaluation metrics such as recall and precision, give a very different result for user satisfaction. This work is similar in nature to the more recent work of Järvelin and Kekäläinen (2000) who presented measures, discounted gain measures, for incorporating differentiated relevance scores into a test collection environment and for incorporating some notion of user satisfaction or user requirements into the evaluation process. De Vries, Kazai, and Lalmas (2004) also modelled novel user-based measures within test collection environments by trying to incorporate user effort, specifically how much effort a user will waste accessing non-relevant sections of an object before accessing relevant material. All these approaches are promising in that they show that simple relevance metrics do not capture user experience of searching and that we can investigate beyond the narrow boundaries of test collections.

It is mainly in the area of evaluation in which the two relevance areas come closest, where the research on understanding the human experience of relevance has a theoretical and practical effect on our provision of search tools. Evaluation measures, such as Järvelin and Kekäläinen (2000), and methodologies, such as Borlund (2003a), demonstrate that there can be a creative interplay between IR and IS.

8. CONCLUSION

The methodology by which we investigate relevance and incorporate relevance into our systems affects how we think about relevance. Classical approaches to IR have taken a definition of relevance that infers a topical relationship between a document and a search request or information need. In a sense, this was not a bad decision as topical relevance tends to subsume other types of relevance and a judgment of non-topical relevance will usually infer a decision of non-relevant (Greisdorf, 2003). However, it is naturally a limited definition of relevance. The fact that this definition of relevance followed through into the test collections we use for evaluation has reinforced this definition within the systems themselves: we evaluate on a system's ability to retrieve topically relevant documents and hence build systems to retrieve topically relevant documents. This strong view of relevance: that is there is a single, objective reality to relevance has been a major obstacle to more naturalistic methods and systems.

Simply redefining this objective reality, e.g. redefining relevant to mean highly relevant, does not necessarily make relevance a more natural concept. How we use relevance in the design of IR systems—what evidence of relevance we see as important, how we believe this evidence should be handled, what inferences we draw from this evidence—define what we see as the task of retrieval systems. It may be the case that we want separate systems for different search purposes, each serving a single rational purpose. Alternatively, we want single systems that incorporate what Saracevic (1996) calls "an interdependent, interacting 'system of relevancies'" (p. 204). We might want, that is, to be able to detect what types of relevance are appropriate at individual points and deal with them appropriately. What we are not able to do at present is to say that there are types of relevance that we should not consider; to say that some relevancies are worthy of further investigation, even if they are more amenable to implementation, and others are not.

The areas of IR covered in this chapter were chosen because either there is a developing core of integrative work between IS and IR relevance research or IR system design would benefit from the incorporation of more of the relatively stable research coming from the IS tradition. The many areas of IR research are not equally affected by relevance; however relevance is the basis of all that we do. This is not a trivial statement as in many pieces of IR research it is difficult to see that there is a recognizably human definition of relevance underpinning the research. This, we believe, is a mistake, as we need to consider where our research fits into the bigger picture of relevance research.

Integrative work between the two research disciplines can strengthen both areas by mapping out research agendas and detailing how these might be supported by new technologies. The more we learn about user searching, the more we can incorporate within IR systems; and the more realistic systems we construct then the more we can uncover about human search behaviour. Not all areas of relevance research make the same assumptions about relevance, and integration does not mean a single solution as there are many possible endpoints from a collaborative approach to relevance. It would be interesting to see where these different collaborations could take us. For example, can we understand better the interaction made by searchers by considering the nature of the interfaces that engender the interaction? Or can we learn what support people require in searching by considering the tasks for which people choose to use IR systems? Or can we establish what level of system complexity is required to adequately respond to different search scenarios? IR systems do not act in isolation of the user's task environment or the user's cognitive environment. Neither are the user's perceptions of relevance neutral to the functionalities offered by the systems which they use. By methodically mapping out the balances and relationships between the searcher and tools they use for searching we can better understand the nature of relevance in searching, the factors that cause relevance to change, and how we can support relevance across the spectrum of information seeking and retrieval research.

REFERENCES

Aalbersberg, I. J. (1992). Incremental Relevance Feedback. *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in IR*, Copenhagen, 11–22.

Barry, C. L. and Schamber, L. (1998). Users' Criteria for Relevance Evaluation: A Cross-situational Comparison. *Information Processing and Management*. 34(2/3), 219–236.

Beaulieu, M. and Jones, S. (1998). Interactive Searching and Interface Issues in the Okapi Best Match Retrieval System. *Interacting with Computers*, 10(3), 237–248.

Borlund, P. (2003a). The IIR Evaluation Model: a Framework for Evaluation of Interactive IR Systems. *Information Research,* 8(3), Paper No. 152. Available at: http://informationr.net/ir/8-3/paper152.html

Borlund, P. (2003b). The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology,* 54(10), 913–925.

Campbell, I. (2000). Interactive Evaluation of the Ostensive Model, Using a New Test-Collection of Images with Multiple Relevance Assessments. *Journal of Information Retrieval,* 2(1), 89–114.

Choi, Y. and Rasumussen, E. M. (2002). Users' Relevance Criteria in Image Retrieval in American History. *Information Processing and Management*, 38(5), 695–726.

De Vries, A. P., Kazai, G., and Lalmas, M. (2004). Tolerance to Irrelevance: A User-effort Oriented Evaluation of Retrieval Systems without Predefined Retrieval Unit. In *Proceedings of RIAO 2004 (Recherche d'Information Assistée par Ordinateur)* (pp. 163–173). Avignon.

Eisenberg, M. and Barry, C. (1988). Order Effects: a Study of the Possible Influence of Presentation Order on User Judgements of Document Relevance. *Journal of the American Society of Information Science,* 39(5), 293–300.

Eisenberg, M. and Hu, X. (1987). Dichotomous Relevance Judgments and the Evaluation of Information Systems. In *Proceedings of the American Society for Information Science 50th Annual Meeting,* (pp. 66–69).

Florance, V. and Marchionini, G. (1995). Information Processing in the Context of Medical Care. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in IR* (pp. 158–163). Seattle.

Greisdorf, H. (2003). Relevance Thresholds: A Multi-Stage Predictive Model of How Users Evaluate Information. *Information Processing and Management*, 39(3), 403–423.

Hersh, W. and Turpin, A. (2001). Why Batch and User Evaluations Do Not Give the Same Results. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in IR* (pp. 225–231). New Orleans.

Harter, S. P. (1996). Variations in Relevance Assessments and the Measurement of Retrieval Effectiveness. *Journal of the American Society for Information Science and Technology,* 47(1), 37–49.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedle, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems,* 22(1), 5–53.

Huang, M.-H. and Wang H.-Y. (2004). The Influence of Document Presentation Order and Number of Documents Judged on Users' Judgements of Relevance. *Journal of the American Society for Information Science and Technology,* 55(11), 970–979.

Ingwersen, P. (2002). *Information Retrieval Interaction*. Taylor Graham.

Janes, J. (1991). The Binary Nature of Continuous Relevance Judgments: A Study of User's Perceptions. *Journal of the American Society for Information Science and Technology,* 42(10), 754–756.

Järvelin, K. and Kekäläinen, J. (2000). IR Evaluation Methods for Retrieving Highly Relevant Documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in IR* (pp. 41–48). Athens.

Katter, R. V. (1968). The Influence of Scale Form on Relevance Judgments. *Information Storage and Retrieval*, 4, 1–11.

Kazai, G., Lalmas, M., and de Vries, A. P. (2004). The Overlap Problem in Content-Oriented XML Retrieval. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 72–79). Sheffield.

Kelly, D. and Belkin, N. J. (2004). Display time as implicit feedback: Understanding task effects. In *Proceedings of the 27th Annual ACM International Conference on Research and Development in Information Retrieval* (pp. 377–384). Sheffield.

Lesk, M. E. and Salton, G. (1969). Relevance Assessments and Retrieval System Evaluation. *Information Storage and Retrieval,* 4, 343–359.

McLaughlin, K. L. and Sonnenwald, D. (2002). User Perspectives on Relevance Criteria: A Comparison Among Relevant, Partially Relevant and Not- Relevant Judgements. *Journal of the American Society for Information Science and Technology,* 53(5), 327–342.

Markkula, M., Tico, M., Sepponen, B., Nirkkonen, K., and Sormunen, E. (2001). A Test Collection for the Evaluation of Content-Based Image Retrieval Algorithms—A User and Task-Based Approach. *Information Retrieval,* 4(3/4), 275–293.

Maron, M. E. (1964). Mechanized Documentation: The Logic Behind a Probabilistic Interpretation. *Statistical Association Methods For Mechanized Documentation*. National Bureau of Standards Miscellaneous Publications, 269, 9–13.

Mizzaro, S. (1997). Relevance: The Whole History. *Journal of the American Society for Information Science and Technology*, 48(9), 810–832.

Mizzaro, S. (1999). Measuring the Agreement Among Relevance Judges. In *Proceedings of the Final Mira Workshop (Mira '99)*. British Computer Society Electronic Workshops in Computing series. Glasgow. 1–13. Available at: http://ewic.bcs.org/conferences/1999/mira99/

Oard, D. W., Mayfield, J., Kharevych, L., Strassel, S., Soergel, D., Doermann, D., Huang, X., Murray, G. C., Wang, J., Ramabhadran, B., Franz, M., and Gustman, S. (2004). Building an IR Test Collection for Spontaneous Conversational Speech. In *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval* (pp. 41–48). Sheffield.

Oddy, R. N. (1981). Laboratory Tests: Automatic Systems. In: *Information Retrieval Experiment.* In: K. Spark Jones (Editor). Butterworths, 156–179.

Reid, J. (2000). A Task-Oriented Non-interactive Evaluation Methodology for IR Systems. *Information Retrieval,* 2(1), 115–129.

Ruthven, I. (2001). *Abduction, Explanation and Relevance Feedback.* Unpublished doctoral dissertation. University of Glasgow.

Ruthven, I., Lalmas, M., and van Rijsbergen, C. J. (2002). Ranking Expansion Terms Using Partial and Ostensive Evidence. In *Proceedings of the 4th International Conference on Conceptions of Library and Information Science (CoLIS 4): Emerging Frameworks* (pp. 199–220). Seattle.

Saracevic, T. (1975). Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science,* 26(6), 321–343.

Saracevic, T. (1996). Relevance Reconsidered. In *Proceedings of the Second Conference on Conceptions of Library and Information Science. Information Science: Integration in Perspectives*. Copenhagen, 201–218.

Salojärvi, J., Kojo, I., Simola, J., and Kaski, S. (2003). Can Relevance be Inferred From Eye Movements in IR? In *Proceedings of the Workshop on Self-Organizing Maps (WSOM'03)* (pp. 261–266). Hibikino.

Sormunen, E. (2002). Liberal Relevance Criteria of TREC—Counting on Negligible Documents? In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in IR* (pp. 324–330). Tampere.

Spink, A. (1996). Study of Interactive Feedback During Mediated Information Retrieval. *Journal of the American Society for Information Science*, 47(8), 603–609.

Spink, A., Greisdorf, H., and Bateman, J. (1998). From Highly Relevant to Not Relevant: Examining Different Regions of Relevance. *Information Processing and Management*, 34(5), 599–621.

Sweeney, S. and Crestani, F. (2003). Supporting Searching on Small Screen Devices Using Summarization. In *Proceedings of Mobile HCI 2003 International Workshop on Mobile and Ubiquitous Information Access*. Lecture Notes in Computer Science, Vol. 2954 (pp. 187–201). Udine.

Tombros, A. and Crestani, F. (2000). Users' Perception of Relevance of Spoken Documents. *Journal of the American Society for Information Science*, 51(9), 929–939.

Tombros, A. Ruthven, I., and Jose, J. M. (2005). How Users Assess Web Pages for Information-seeking. *Journal of the American Society for Information Science and Technology,* 56(4), 327–344.

Tiamiyu, M. A. and Ajiferuke, I. Y. (1988). A Total Relevance and Document Interaction Effects Model for the Evaluation of IR Processes. *Information Processing and Management*, 24(4), 391–404.

Vakkari, P. (2000). Cognition and Changes of Search Terms and Tactics During Task Performance: a Longitudinal Study. In *Proceedings of RIAO 2004 (Recherche d'Information Assistée par Ordinateur)* (pp. 894–907). Paris.

Vakkari, P. and Sormunen, E. (2004). The Influence of Relevance Levels on the Effectiveness of Interactive IR. *Journal of the American Society for Information Science and Technology*, 55(11), 963–969.

Voorhees, E. M. (2000). Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness. *Information Processing and Management*, 36, 697–716.

Voorhess, E. M. (2001). Evaluation by Highly Relevant Documents. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 74–82). New Orleans.

CHAPTER 5

NEW COGNITIVE DIRECTIONS

NIGEL FORD
University of Sheffield

1. INTRODUCTION

This chapter explores cognitive issues relating not only to the use and effectiveness of information retrieval (IR) systems, but also to the generation of new research knowledge relevant to enhancing such use and to the development of improved systems. There is still much that we do not know about how effectively, or not as the case may be, current IR systems support different individuals and groups of people in effectively acquiring the knowledge they need to help them in the tasks in which they are engaged. We need greater levels of knowledge of how we might optimize the symbiotic interactions between IR systems and people working in different contexts, and of the directions we might most profitably take in the development of new systems.

The chapter begins by establishing working definitions of “knowledge”, “information”, “understanding”, and related concepts including “information retrieval” and “information behavior”. Such definitions are necessary if we are to achieve any clear focus on what are important issues for the evaluation and development of IR systems and practices. *Knowledge* is defined here as a conceptual system that enables its possessor (whether human or machine) to act in the world (e.g. in the form of autonomous activity or response to stimuli). *Information* is defined here as potential knowledge, and *understanding* as the process of transforming information into knowledge. Perceiving an information need represents the realization that one is in a situation for which one’s existing knowledge is not adequate. Information is needed to generate, via understanding, appropriate new knowledge.

However, the *need for knowledge* is broader than the *information need* in that the former entails defining the nature of the sought-for knowledge, as opposed to the type and nature of the information which may be used to generate it. Importantly in this context, functionally the same knowledge may be generated via the processing of alternative items of information. The realization and specification of the knowledge need, along with the evaluation of information behavior in terms of its contribution to satisfying the knowledge need, are termed here *knowledge behavior*.

Information behavior comprises those activities entailed in the acquisition of information that may be transformed into knowledge. It subsumes *information encountering* and *information seeking*, the latter subsuming *searching* and *retrieval*. Although *information retrieval* is often used narrowly to denote the latter activity taking place within information seeking, a broader definition is used here to include the development, use and evaluation of systems designed to support the range of activities denoted by the term information behavior.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 81–96.

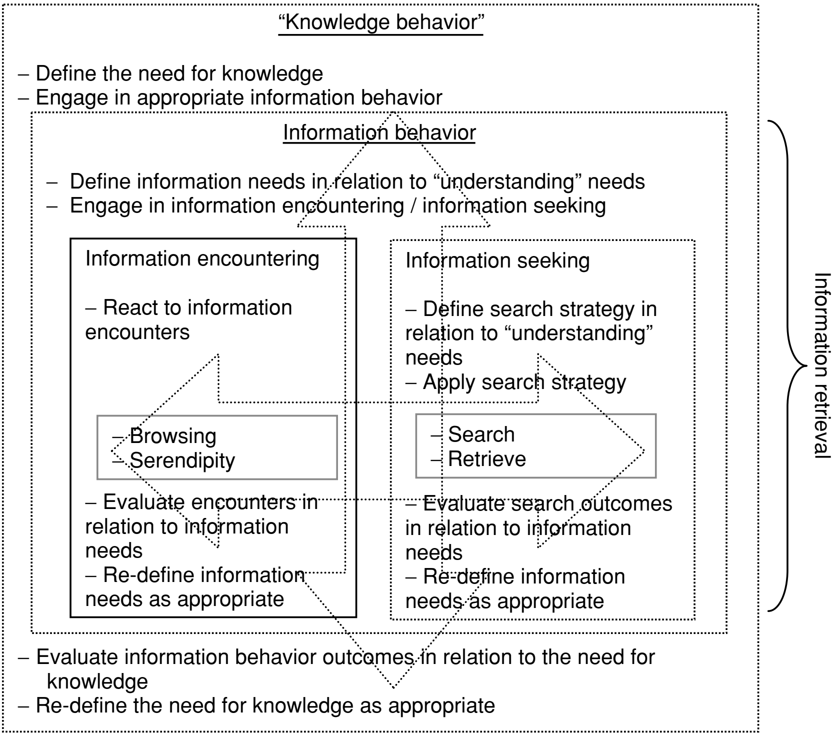

Figure 1: Information retrieval as defined here

These concepts are depicted in Figure 1. The dotted arrows signify that, as argued below, there may be fluid movement of intellectual and behavioral activity across the areas indicated.

Research is defined here as the generation of knowledge via systematic inquiry (e.g. to explain and/or predict some phenomenon). *Information retrieval research* is defined here as the generation of information designed to be helpful for the development, use and evaluation of systems for supporting information behavior.

Cognitive aspects of IR are defined here to encompass those aspects of the development, use and evaluation of information behavior support systems that explicitly reference the thought processes entailed in that behavior. "Cognitive" is often used, particularly in education, to denote logical thought as opposed to attitudes and feelings which are referred to as "affective". No such distinction is observed here, and "cognitive" is used in its broader sense to refer to both logical and affective thought processes.

2. THE "DARKNESS TO LIGHT" RATIO

Inability to move beyond the relatively modest recall/precision plateau of IR system performance, and failure of user models apparently to have much to say to information system designers would seem to indicate that the processes whereby IR systems might optimally contribute to effective information behavior are by no means fully understood.

Ford (2000a) has characterized our lack of such knowledge in terms of a “darkness to light ratio”:

> Complex IR systems have been and are being built—but they are often based on relatively untested theoretical assumptions about users, not on any valid and robust empirically-based understanding of them. Developments in IR system design increase strategic options open to the user—but tell us nothing of how such options may be deployed to achieve optimally effective searching . . . Our desire to develop practically useful knowledge may encourage us to be over-optimistic in relation to the generalizability of what are essentially very sporadic research efforts. This over-optimism may mask the scale of the darkness to light ratio and the fact that we are falling very short in terms of the critical mass required to support the generalizable knowledge we seek (Ford, 2000a, p. 626).

Such an observation would suggest that little has changed since Dervin and Nilan’s (1986) seminal paper highlighting not only the lack of symbiosis between user-based and systems-based research. However, the apparent incompatibility of the predominant paradigms adopted by each. Although, as acknowledged by Johnstone, Bonner, and Tate (2004), alternative systems paradigms such as Checkland’s SSM do allow integration between the two.

At one level user studies have indeed been useful in informing system design. Knowledge, for example, that clinicians regard randomized control trials (RCTs) as a gold standard of quality information relating to a particular class of clinical problem has driven the development of health-related IR facilities incorporating search filters that identify RCT-based studies. User consultations establishing the usefulness of high quality systematic reviews has driven the development of health-related IR systems that rank items retrieved from quality sources such as the Cochraine Collection more highly than sources of lesser consensually agreed pedigree.

However, the incidence of such significant influences of user studies on major aspects of *general* system design is small. Human-computer interaction (HCI) studies have been useful in enabling system designers to address issues relating to phenomena such as user preferences and cognitive ergonomics. But there is a real sense in which user and HCI studies have very little to say to designers in relation to more substantive parameters of system design that affect the effectiveness of information behavior.

Järvelin and Ingwersen (2004), for example, note the relative exclusion (a) from traditional IR evaluation studies, of consideration of users in the context of their work tasks, and (b) from information seeking studies, of consideration of IR systems and features. In consequence, we are largely ignorant of how effective or ineffective IR systems are in different contexts.

Insofar as they do say anything to IR system designers, a number of recent models of aspects of human information behavior emphasize the complexity of information seeking and the fuzziness of notions of “relevance” and “retrieval effectiveness”. These flag the extent to which, beyond a certain level of complexity in problem-solving and information seeking, there is no easy mapping of user characteristics onto those parameters of IR system design that may render information behavior effective.

To extend Ford’s “darkness and light” analogy, we may recall that astronomers, faced with a mismatch between explanatory theory and the observed behavior of galaxies, inferred that either gravity does not work in the ways we think it does when scaled up to galactic proportions, or that some “dark matter” must be brought into the equation. More recent observations would also seem to indicate the presence of some further unaccounted-for “dark energy”—the origin and characteristics of which, like dark matter, are unknown.

Might we also need to take into account some dark un-comprehended factors before we can more fully understand human aspects of IR? Yet what might be—and where might reside—the dark unknowns that such understanding might require us to discover? It is argued below that we are at least beginning to make inroads to such discovery.

3. COMPLEXITY IN INFORMATION BEHAVIOR

Since the 1980s, there has been a substantial body of user-related research that has adopted what Dervin and Nilan (1986) termed an “alternative paradigm”. The knowledge deriving from such studies paints a picture of human information processing that seems far removed from the “traditional” information systems view of information processing as characterized by order, sequence, analysis, objectivity and control.

Increasingly, studies are bringing into question the ecological validity of views characterizing human information behavior in terms of a relatively ordered sequence of activities. Foster (2004), for example, in his doctoral study of inter-disciplinary researchers, notes the non-linear nature of the interactions between the processes entailed in information seeking. In his model, processes and contexts shift and interact in complex, dynamic and subtle ways. The non-linear and holistic uses of, and interactions between these processes are “analogous to an information seeker holding a palette of information behavior opportunities, with the whole palette available at any given moment”. Even such an apparently discrete stage as “problem definition” was not clear cut, problems being constantly redefined throughout information-seeking.

Based on a review of the educational and cognitive research literature, Ford (2004c) presents a model of information needs and relevance judgments which echoes Foster’s model insofar as it views these aspects of information behavior as sometimes characterized less by discrete sequences than by the relatively holistic processing of many interacting factors. He notes that information needs and relevance judgments are influenced not only by the nature of the task, but by a range of factors such as motivation, anxiety, and levels of existing knowledge and experience. The profiles of information seekers in terms of these factors may be thought of as individually different, and time-dependent, templates through which the individual information seeker perceives her or his “information needs” and judges “relevance” throughout the information seeking process. The templates change with the acquisition of new information, and “relevance” may be viewed very differently via changed templates as information seeking progresses.

Aspects of this picture relating to the dynamic, shifting nature of relevance judgments are echoed in an in-depth longitudinal ethnographic study of academic researchers reported by Anderson (2005), who also stresses the *context-dependency* of, and the role of *intuition* in, the making of such judgments. According to Anderson, observed experiences are shifting and dynamic. There relevance judgments must be ongoing activities inserted into seeking and searching processes. The user uses prior knowledge, experiences plus intuition in these judgments. To understand relevance judgments therefore means understanding this context.

The ongoing assessment of information needs and the relevance of retrieved information may at least sometimes be characterized more in terms of the holistic processing entailed in the reconfiguring of a gestalt than the more analytic step-by-step working through of logical procedures. Such processing, linked with the notion of intuition as noted by Anderson, is reminiscent of the "flash of insight" often reported as characterizing divergent (creative) as opposed to more convergent thinking (Ford, 1999a):

> It would also appear that to a significant degree, creativity often depends on the relatively unplanned and often sudden recognition of a similarity between disparate entities—as opposed to the more progressive, controlled jettisoning of dissimilarity. This is characteristic of . . . perceptual, holistic, parallel information processing discussed above. Often descriptions of instances of creative thought suggest a sudden perception or realization—a flash of insight when things come together—occurring when the person is not intensely focused on the particular problem (Ford, 1999a, p. 532).

Convergent information processing is an essential aspect of information retrieval. Indeed, at one level, it seems eminently sensible to try to establish maximum convergence between concepts specified by an information seeker in a query to a retrieval system and concepts contained in documents retrieved. Interestingly, however, Spink and Greisdorf (1997) found an inverse relationship between high levels of relevance—defined in terms of the extent to which retrieved documents matched user queries—and the generation of new ideas and directions by researchers engaged in online searches:

> 'highly' relevant items do not change the users' cognitive or information space in relation to their information problem . . . Highly relevant items may not relate to a shift in a user's information problem towards resolution but reinforce the current state of the user's information problem and knowledge state . . . Items retrieved that are not 'highly' relevant, but partially relevant, are related to shifts in the users' thinking about their information problem by providing new information that leads the users in new directions . . . (Spink and Greisdorf, 1997, p. 276).

"Highly relevant" items contributed to relatively convergent thinking—helping users confirm what they already knew and to maintain existing directions, as opposed to relatively divergent, or creative thinking—generating new ideas and directions.

This is perhaps not surprising in that IR systems work essentially via convergent processing to achieve as close a match as possible between queries and documents. However, the complementary role of divergent information processing is stressed by Ford (1999a), who claims that high levels of creativity depend on the type and level of dissimilarity applying to the context and/or the nature of the discrete entities brought together and integrated by a similarity relationship. He argues that it is the extent of dissimilarity, and the level of abstraction, that differentiate creative from less creative forms of thinking.

Divergent information processing entails a relaxation in the specification of precisely what information is required when searching. Indeed, Strauss and Corbin (1997) note how insights often occur when people switch off from conscious thinking about a problem—which can actually hinder the process of realizing "the sudden hunch, the flash of insight, the brilliant idea, or the profoundly different theoretical formulation".

Such relaxation and implicit openness to new ideas can sometimes be well served by serendipitous encounters with what may turn out to be useful—though unanticipated—information. Foster found evidence of serendipity as a widely experienced, and valued, aspect of information behavior amongst his sample of interdisciplinary researchers. Serendipity itself has often been associated with the "flash of insight", previously referred to in relation to divergent thought, characteristic of many accounts of creativity. Foster and Ford (2003) note the difficulties entailed in researching such an essentially "fuzzy" concept as serendipity, since it is by definition not particularly susceptible to systematic control and prediction. Nevertheless, serendipity did emerge in Foster's study as an important aspect of researchers' information behavior and their generation of new ideas.

4. RESEARCH APPROACHES

The small selection of models presented in the previous section is indicative of a wide range of studies that have adopted research perspectives that are very different from the "traditional" information systems perspective criticized by Dervin and Nilan in 1986. They are examples of research that has adopted a predominantly "right column" research perspective, as opposed to the relatively "left column" perspective often associated with more traditional IR studies, shown in Table 1. This table presents a polarized view of characteristics associated with key dimension of difference between common research approaches, representing a blend of aspects presented by Dervin and Nilan (1986) and Ford (1999b).

At one level it seems sensible to concentrate on those parts of phenomena we can most readily understand, and that offer us the clearest prospect of establishing cumulatively useful building blocks of understanding. The essentially statistical research paradigm represented by the left column would seem to offer a particularly

Table 1: Differences characterizing common research approaches

Quantitative, statistical	Qualitative, interpretative
Atomistic (focusing: component parts)	Holistic (focusing: whole)
Sequential processing	Parallel processing
Isolate variables	Preserve complexity life-as-situations
Precise analytic defn. + measuring	"Fuzzy" intuitive concepts
External observed behavior	Internal cognition
Study discrete relationships	Study complex interacting relationships
Concepts	Sensitizing concepts
Logic-mathematical	Intuitive-social
Convergent processing	Divergent processing
Control	Relaxation of precise control
Predictability	Idiosyncrasy
Objectivity	Subjectivity
Mechanisitic	Constructivist
Trans-situational, generalizability	Context bound transferable "insights"

cumulative-friendly type of evidence with its emphasis on objectivity and measurable generalizability, and understandably much IR research has adopted such an approach.

This is fine, so long as we remember that it paints only a part of the picture. Human activity, including interactions between humans and IR systems, is study-able only to a certain extent from such a perspective. As perceived by many researchers who have pursued and developed Dervin and Nilan's (1986) "alternative paradigm", the softer multi-angle lighting it can provide, admittedly with its many optical tricks and shadows, is at least a necessary complement to the intensely bright and focused, but somewhat harsh and linear searchlight provided by its polarized Table 1 alternative in illuminating some of the dark unknowns associated with the human aspects of IR that elude us.

Different research perspectives and approaches can generate fundamentally different kinds of knowledge, each with its distinctive strengths—and weaknesses. Indeed, it may be that the distinctions to some extent reflect fundamental biological distinctions within individuals. Research in neurological science has characterized the left hemisphere of the brain as (Raina, 1979, p. 10):

> ... a rational-linear mind specializing in sequential processing logical analytical thinking and verbalization. ... This is the mind that requires structure and order, which processes perception and sensory input in logical and linear modes.

Whereas the right hemisphere (Samples, 1977, p. 688) "houses spatial perception, holistic understanding, perceptual insight ... visualization and intuitive ability. Its mode is metaphoric, analogic and holistic. ... This side of the brain thrives on multiple relationships processed simultaneously."

It has been argued elsewhere (Ford, 1999b; 2000a) that the two broad perspectives outlined in Table 1 map onto two different, but equally essential basic components of human understanding identified and elaborated in studies over more than 30 years by the eminent cognitive scientist and cybernetician Gordon Pask, namely: procedure-building and description-building.

Procedure-building relates to mastering procedural details—the step-by-step logical evidence and arguments supporting the "big picture". Description-building entails the construction of that big picture—the holistic conceptual map of concept interrelationships. As noted by Entwistle (1981, p. 93): "Pask has likened these two aspects to the way an architect designs a building. He has to build up the overall plan (description building) and also to work out the detailed processes, and the logistics of those processes (... procedure building) whereby the plan can be converted into an actual building."

If research paradigms are the products of human cognition, then it should hardly seem surprising if the distinctions discussed reflect those identified in research into cognition itself. However, these phenomena may also share sources of potential weakness. Indeed, Pask found that individuals may be biased towards one aspect more than another, a pronounced bias resulting in the individual stopping short of full understanding. For example, the individual strongly biased towards description-building may stop short of a detailed examination of the procedural evidence supporting the bigger picture. In doing so, he or she risks over-generalization leading to a concept-relationship map of the phenomenon in question that is insufficiently supported with detail and evidence and thus potentially invalid. The converse risk equates, basically, to failing to see the wood for the trees, with very well developed and supported knowledge of

Table 2: Description-building, procedure-building and versatility

High		Procedure building		Low
Narrow and deep but fragmented Knowledge lacking context	Thorough grasp of logic and detail	Versatility: Strong in both conceptual interrelationships and supporting detail and logic	Good grasp of How broad concepts interrelated	Broad contextual knowledge But lacking supporting evidence
Low		Description Building		High

parts—but lacking the integration necessary to generate a coherent map of how the parts inter-relate to form the whole:

- "Any weakness either in the plan, or in . . . the operations and procedures which will satisfy the implicit demands of that plan, will prevent the building being satisfactorily completed (understanding being reached)." (Entwistle, 1981, pp. 93–94)
- Some individuals display a sufficiently strong bias to display the weaknesses described above. However, "versatile" individuals are able to combine both description-building and procedure-building to achieve deep understanding (Pask, 1975, 1976a, b, 1988; Ford, 2000b).

Table 2 depicts these complementary components and the weaknesses associated with different types of imbalance.

Ford (1999b) has argued that this operates not only at the level of individuals, but also at the level of entire fields of activity and study. But how might the stakeholders in IR—system developers, researchers and users—achieve understanding sufficiently deep to bring about effective knowledge and information behavior? Who needs what knowledge? And in particular, to what extent do IR users need to develop deep understanding, rather than expect intelligent IR systems somehow to shield them from such a requirement?

4.1. System Control

With developments in ambient computing, recommender software, agents and IR systems that engage in anticipatory background retrieval, the range of intellectual activity being performed *by* systems *for* the user as opposed to directly by the user is increasing, despite the use by such systems of user feedback. Developments in the semantic web enabling machines to reason across ontologies also offer the prospect of the inferences entailed in intelligent reasoning being increasingly made *for* as well as *by* users.

Where the needs of users are clear and simple, and in cases where inferences are relatively factual rather than value-laden or potentially contentious, it will be possible for machines to takeover from users the drudgery of time-consuming and intensive processing. In other words, *supplanting* cognitive activity—seems eminently sensible and, conceptually, relatively unproblematic. However, in the context of the complexity, subjectivity and content-dependency characterizing information needs and the criteria determining their satisfaction described in section 1.3, the notion of supplantation becomes more problematic.

4.2. User Control

The notion of information needs and their satisfaction as to a significant extent multi-faceted, dynamic and context-bound phenomena may often be at odds with the relatively simple—but objectively measurable and reliable—conceptions of relevance and need typical of many quantitative and reductionist studies which seek to establish objectively definable and reliable, context-independent characteristics and performance of IR systems and their components.

If information needs and relevance judgments were in reality always so simple, then it might be relatively straightforward for machine systems to supplant the human processes of selecting and applying appropriate IR components in relation to particular information needs. But in reality they may often be more complex, fuzzy and problematic than is taken into account in many IR experiments, which must over-simplify them in order to characterize their behavior in objective and generalizable ways. In cases of high complexity, it might be better if users themselves were able to select and deploy IR components to suit their potentially highly individualized—and dynamically changing—needs and circumstances.

Putting users in direct control might be a way of reducing the semantic gap between (a) their information needs and relevance criteria, and (b) IR systems' understanding of these needs and criteria. However, users would need to possess appropriate knowledge of the alternative IR components available to them (for example, retrieval algorithms, output options, query expansion facilities, inference rules, learning mechanisms, etc.), and of the implications of choices, sequences and combinations of them in terms of system behavior and performance in relation to their needs. The following section explores issues relating to the acquisition by users of such knowledge.

5. USERS' ACQUISITION AND APPLICATION OF IR-RELEVANT KNOWLEDGE

Users might acquire such knowledge in at least two ways. One route would be to gain access to information about how particular IR components behave, which is as objective and generalizable as possible so that they could make an informed choice with reasonable effort. If information about the characteristics and performance of a particular component is objective and generalized, then they may feel reassured that it is likely that these characteristics and performance will be delivered in the context of their particular search.

Ironically however, as argued in Section 1.4, the very objectivity and generalizability that render such information so context-independently applicable also tend to limit the ecological validity of that information when applied to complex contexts, due to the simplification and isolation of variables required to facilitate the necessary statistical analyses.

To some extent IR components *will* behave in predictable ways, and their characteristics and performance *will* be objectively describable. However, there are likely to be other dimensions of effect that are not susceptible to objective measurement and the identification of reliable generalizable performance. Even were the characteristics and performance of individual components definable in objective and context-independent terms, they may not behave in the same way when operating in the complex interactions that may characterize the processes of attempting to satisfy less simple information needs.

Furthermore, even if such interactions were to display predictable context-independent characteristics and performance, it would be difficult for a machine system optimally to select, sequence, combine—and dynamically change such choices, sequences and combinations—in such a way as to "second guess" the individual user—assuming for the sake of argument that s/he were fully cognizant of the choices available and of their behavioral and performance implications—in the context of highly complex and dynamically changing conceptions of need and relevance.

A second approach whereby the IR user could acquire appropriate knowledge would be to gain access to context-rich information, high in ecological validity—for example, case-study material that might map well onto the complexities and subtleties of his or her particular needs and context. In qualitative research terminology, such information might be *transferable* rather than *generalizable*. Transferability is a "resonance" between contexts whereby knowledge gained in one context may illuminate other contexts insofar as the new context shares important features with the original. The potential for transferability is dependent on the description of the original context being sufficiently rich in detail to allow its resonance with the new one to be assessed in terms of how similar it is and what features it shares. Transferability depends both on rich knowledge of the new context (the user's knowledge need and associated information need and relevance criteria), and on the rich description of the original research context. Such description must be sufficiently rich to enable a mapping of similarities between the two.

However, the very individuality of complex needs and relevance criteria poses difficulties for transferring knowledge in this way, since highly context-specific knowledge is only transferable to other contexts if there is a rich mapping of elements. Unless descriptions of many complex scenarios were stored and available it may be difficult to locate information capable of being transformed into knowledge potentially transferable to the new situation.

Rich contextual description entails complex, multifaceted sets of elements, and it is always possible, in relation to a complex information need experienced by a particular individual in a particular context, that important ones have not been included amongst those that have been explicitly described and mapped. Also, the more novel and unique the information need, the more difficult it would be to locate recorded cases that share a richly similar nature and context. Finally, even were such cases located, determining the extent of their transferability to the new context would be likely to require in-depth, critical—and time-consuming—reading of rich contextual descriptions.

The same problem applies to system design, as noted by Järvelin and Ingwersen (2004), who state that a counter-argument is that with so many information seeking contexts and systems, design/evaluation of IR systems is unmanageable, making it a tall order to use the indicated design and evaluation approach, so therefore keep the traditional design and evaluation approach. However, they go on to conclude that a better basis for design than the de-contextualized traditional assumptions/measurements is to find the typical strengths, weaknesses and uses of systems, as well as user perceptions. This view, however, implies that appropriate information concerning the characteristics and performance of IR components is generated *by* researchers *for* users.

A third route to the acquisition of appropriate knowledge is for IR users themselves to generate, or at least be involved in the generation of, knowledge relevant to their own particular needs and circumstances. One such route could entail a form of action research—the "action" being the development and practice of effective knowledge and information behavior using IR systems. Action research is capable of explicitly taking advantage of subjectivity, and of the contingent dependencies that may arise from interactions between observers and observed that, if not made explicit, can represent a characteristic weakness of qualitative research paradigms. As Ford (2004a) notes, contingent dependencies need not be inadvertent, and can represent intended outcomes of a planned intervention. When made explicit, such interactions can include negotiations relating to what might be desirable outcomes, and how they might be achieved through the research.

In the field of education, Levy (2003) provides a succinct description of action research, and its potential for generating personally—and practically—relevant knowledge:

> Recognizing the uniqueness and complexity of specific educational situations, action researchers assume that professional competence is more than just a matter of routinely applying universal rules of practice or purely technical know-how prescribed by some external source of guidance. Instead, competence is taken to be a matter of being able to perceive, and take forward, the right course of action when in a particular situation . . . it is based on knowledge that is embedded in personal experience and that is essentially concrete, context-specific, and open to re-interpretation (Levy, 2003, p. 90).

This generally involves dialogue with other stakeholders in the phenomenon being researched. Levy goes on to explain that educational action research is based on interpretive/reflective practice, that is investigated and evaluated inductively, with a new perspective on assumptions, values etc. relating to program participant perspectives.

However, the essential activities of critical reflection and the challenging of one's own assumptions, values, aims and actions may also be achieved via a "critical friend", or indeed via systematic critical "inner dialogues" such as those inherent in Pask's *Conversation Theory*. Ford (2004b) states:

> Theory generation and testing via the exchange, explanation and justification of mental models is the essence of conversations. Although particularly applicable to information processing situations entailing teachers and learners, the theory is also applicable to less formal and mediated information use. Pask points out that conversations between different p-individuals can take place within the brain of a single person—representing "inner dialogues" as people come to understand (Scott, 2001) (Ford, 2004b, p. 773).

The "p-individuals" referred to in the quotation above are conceptual systems—for example, particular understandings of topics. Such understandings (which may include particular points of view) can be tested out—for example, compared, contrasted and evaluated in terms of the presence and absence of evidence supporting them—in what Pask terms the "conversations" central to his *Conversation Theory* (Pask, 1975, 1990; Pask and Gregory, 1986; Pangaro, 1997; Kybernetes, 2001). These can take place within an individual, who may test out different competing understandings in his or her own mind via the "inner dialogues" referred to—or in conversational interactions with others.

Viewed from this perspective, the development of effective information behavior (including information seeking and IR, as shown in Figure 1) is a process of developing personally relevant and personally validated knowledge via experimentation and critical reflection—whether individually or with others.

The type of knowledge that may be generated from such activity is well described by Levy (2003) in the context of action research:

> [A]ction research therefore does not aspire to the construction of propositional knowledge (*episteme*), or to the production of purely technical know-how. Instead, the aim is to construct validated, 'practical' knowledge (the Aristotelian concept of *phronesis*) through critical analysis of specific . . . situations within the context of action—that is, through *praxis* . . . This is seen as a dialectical process that is constitutive of effective action *and* new knowledge . . . (Levy, 2003, p. 90).

However, Levy goes on to explain that the applicability of such personally relevant and validated knowledge need not be restricted to the individual. Theory generated through praxis—sometimes referred to as "living theory"—is shareable within professional communities, who can in this way build up a shared knowledge base.

Ultimately, we must all build the personal knowledge structures that drive and guide our behavior, and in doing so must determine the nature and extent of the evidence we deem sufficient to inform our actions. In the context of complex information needs, this is likely to entail not only processing information generated by significant others such as IR researchers and system designers, but also the generation of personally relevant and validated knowledge generated through one's own experimentation and experience—whether or not moderated by some significant other(s).

In the context of (a) complex, dynamically evolving information needs and relevance criteria, and (b) the availability to the information seeker of choice of and/or control over complex and multifaceted IR systems and components, the level of knowledge required to make informed choices and to exercise appropriate control may be such that, to some extent, users need to become their own practitioner-researchers.

6. THE FUTURE

We are fairly good at understanding weather systems such that we can predict with reasonable accuracy weather patterns in the short term. However, the complexity of variables and interactions entailed when we attempt to understand the way in which such systems behave over longer time periods escalates rapidly.

Understanding the generation and satisfaction of sophisticated information needs may entail a level of complexity more analogous to that associated with long term rather than short term weather systems. In our attempts to understand weather systems, we need to develop and test hugely complex computer-based models entailing large numbers of variables and complex interactions. It may be that efforts of a similar scale and complexity are needed for us to advance our understanding of those cognitive aspects of IR that elude us.

The most significant illumination in our understanding of these unknowns may in the future derive from more powerful and sophisticated quantitative models. However, it is also possible that such approaches may take us only to a certain plateau, and that interaction between different perspectives and research approaches will be required to make further inroads. Whilst complex physical phenomena such as weather may ultimately be understandable via quantified models—albeit of a complexity not yet achieved—social and psychological phenomena, including knowledge and information behavior, may ultimately not be fully understandable in these terms.

We have a great multiplicity of information generated from different perspectives (system designers, researchers and users) and based on different levels and types of knowledge. We arguably need some sort of dynamic, interactive, multi-perspective and flexible "evidence map" that can enable and record the integration, cumulation, and disputation of—and other forms of interaction between—such disparate information sources. Ford (2000a) has noted such a need in relation to researchers, although the same, arguably, applies to users:

> Researchers from a range of different disciplines . . . need to be able to find relevant IR research findings in response to searches formulated in terms of their own particular problems and perspectives. Such multi- and inter-disciplinary searches would benefit from a range of ontologies designed to map such requests onto more specialized IR terminology. However, as well as a range of heterogeneous reviews and ontological mappings of search terminologies, we also need multi-perspective and multi-purpose conceptual mappings more complex than those provided by ontologies (Ford, 2000a, pp. 634–635).

Developing such an evidence map—if it is to possess the ability to enable and map different perspectives and interactions—will be a non-trivial task. Systematic reviews and meta-analyses to some extent provide an integrated view of particular areas of research, but entail imposing a particular filter, structure and perspective. Ideally, we need something more flexible, multi-faceted, dynamic and interactive, capable of responding to questions of the type:

> *"Let me explore topic X in terms of aspect Y, from perspective Z. I want to keep X and Y exactly as I have defined them, but would like to explore variations in Y".*

Although relatively undeveloped at the present time, it might be worth exploring the potential, as a basis for such a system, of ZigZag—Ted Nelson's (1998, 1999, 2001, 2004) new hypertextual paradigm. According to Moore and Brailsford (2004), the ZigZag system renders the complex comprehensible via uncovering underlying information structure. Storing and visualizing information interconnectedness are better provided for than traditional hypertextual representation, via hyperstructuring to show relationships in different lines for any scaffolding (implicit/explicit) being used.

Moore and Brailsford go on to describe exploratory work in relation to bioinformatics, which they consider to be characterized by unusually high levels of intrinsic complexity. They comment on the ability of ZigZag to render complex structures more easily browsable and comprehensible.

Nelson (2004) describes the basic approach as starting from (1) orthogonally connected data items (zzcells) plus (2) untyped connections (zzlinks) to (3) construct a cross-connected fabric of data (zzstructure) which is easy to make visual, interactive plus it is easily programmable. Criss-crossed cell lists are given dimensions, along which cells can be viewed, crossed, plus they are amenable to operations. You can view zzstructures in 3D (but with option to view more than three dimensions because cells exist in all dimensions). On the other hand, each new dimension welcomes every cell when the new dimension is constructed, and links can be constructed along the new dimension to any cell when it is created, to any cell in the system.

Two of the principal features of zzstructure are the flexibility with which alternative views can be obtained of complex multi-faceted data, and the ease with which new perspectives can be added, in the form of new dimensions, without the need to reconfigure information already stored. These features may be extremely helpful in enabling us to explore information—to navigate, for example, a complex map of concepts and relationships representing research findings—from alternative perspectives and with different questions in mind, and to change and develop our perspectives and questioning as we explore.

However, it would also be desirable to have access to a system that, as well as enabling such flexible exploration could facilitate *interaction* between such multi-faceted and multi-perspective information. This interaction would entail processes of comparison, identification of conflicts and negotiation of shared agreements. In this way the evidence base could be refined and extended by the generation of modified, confirmed, rejected and extended concept structures at different levels and from different perspectives—both individual and group, including system developers, researchers and users.

Based on Pask's *Conversation Theory* and on preliminary developmental work by Pask, Pangaro and others (Pask, 1975, 1990; Pask and Gregory, 1986; Pangaro, 1997; Kybernetes, 2001), Ford (2005) has proposed such a system in the form of a "conversational information system". The term "conversation" is used here as rigorously defined by Pask in *Conversation Theory*. Conversations essentially entail the type of comparison, identification of conflicts and negotiation of shared agreements referred to above. *Conversation Theory* is well suited to handling interactions and negotiations between different perspectives.

The concept representation developed by Pask and utilized in *Conversation Theory* consists of "entailment structures". These are capable not only of representing the conceptual structures (understandings) of different individuals, but also of enabling the checking of such structures for coherence, since the properties of a well-formed entailment structure enable the automatic detection of ambiguities, gaps and contradictions as new knowledge is added. They can also facilitate the comparison of different understandings, and the identification of agreements and disagreements between participants. This allows the representation of different but equally valid understandings constituting different perspectives, and individuals' different constructed "realities".

Prototypes of Nelson's ZigZag system have been developed and are being tested (e.g., Moore and Brailsford 2004), and there has also been much work developing prototype Paskian conversational systems based on entailment structures (e.g., Pangaro, 1997; Heylighen, 2001; Zimmer, 2001). However, it is not clear how conversational systems might interact with zzstructure data representations. As noted above, the former are based on entailment structures—a particular type of concept representation the characteristics and qualities of which enable and drive conversational interactions. It may be that entailment meshes could represent particular views of—thus operating at a higher level of abstraction than—structures, which could form the basic level of representation.

Much work remains to be done at both conceptual and practical system development levels. Such a vision essentially represents, at this point, what Pask has termed "description-building" (described in section 1.4), and this would need to be complemented by appropriate procedure-building were the ideas to be successfully implemented. But if successfully developed, such an approach could represent new cognitive directions for the development of a new type of interactive evidence map relating to IR—itself representing a novel form of IR to complement more traditional forms of support for the practice and development of knowledge and information behavior.

REFERENCES

Anderson, T. D. (2005). Relevance as Process: Judgements in the Context of Scholarly Research. *Information Research*, 10(2) paper 226. [Available at http://InformationR.net/ir/10-2/paper226.html]

Dervin, B. and Nilan, M. (1986). Information Needs and Uses. *Annual Review of Information Science and Technology*, 21, 3–33.

Entwistle, N. J. (1981). *Styles of Learning and Teaching*. London: Wiley.

Ford, N. (1999a). Information Retrieval and Creativity: Towards Support for the Original Thinker. *Journal of Documentation,* 55(5), 528–542.

Ford, N. (1999b). The Growth of Understanding in Information Science: Towards a Developmental Model. *Journal of the American Society for Information Science*, 50(12), 1141–1152.

Ford, N. (2000a). Improving the 'Darkness to Light' Ratio in User-Related Information Retrieval Research. *Journal of Documentation*, 56(6), 624–643.

Ford, N. (2000b). The Increasing Relevance of Pask's Work to Modern Information Seeking and Use. *Kybernetes*, 30(5/6), 603–629.

Ford, N. (2004a). Creativity and Convergence in Information Science Research: The Roles of Objectivity and Subjectivity, Constraint, and Control. *Journal of the American Society for Information Science and Technology*, 55(13), 1169–1182.

Ford, N. (2004b). Modeling Cognitive Processes in Information Seeking: From Popper to Pask. *Journal of the American Society for Information Science and Technology*, 55(9), 769–782.

Ford, N. (2004c). Towards a Model of Learning for Educational Informatics. *Journal of Documentation*, 60(2), 183–225.

Ford, N. (2005). Conversational Information Systems: Extending Educational Informatics Support for the Web-Based Learner. *Journal of Documentation*, 61(3), 362–384.

Foster, A. (2004). A Nonlinear Model of Information-Seeking Behavior. *Journal of the American Society for Information Science and Technology*, 55(3), 228– 237.

Foster, A. and Ford, N. (2003). Serendipity and Information Seeking: An Empirical Study. *Journal of Documentation,* 59(3), 321–340.

Heylighen, F. (2001). Bootstrapping Knowledge Representations: Via Entailment Meshes Via Semantic Nets to Learning Webs. *Kybernetes,* 30(5/6), 691–722.

Järvelin, K. and Ingwersen, P. (2004). Information Seeking Research Needs Extension towards Tasks And Technology. *Information Research,* 10(1) paper 212 [Available at http://InformationR.net/ir/10-1/paper212.html]

Johnstone, D., Bonner, M., and Tate, M. (2004). Bringing Human Information Behaviour into Information Systems Research: An Application of Systems Modelling. *Information Research,* 9(4) paper 191. [Available at http://InformationR.net/ir/9-4/paper191.html]

Kybernetes (2001). Vol. 30 Nos. 5/6 and 7/8 are devoted to Pask's work.

Levy, P. (2003). A Methodological Framework for Practice-Based Research in Networked Learning. *Instructional Science,* 31(1–2), 87–109.

Moore, A. and Brailsford, T. (2004). Unified Hyperstructures for Bioinformatics: Escaping the Application Prison. J*ournal of Digital Information,* 5(1). Article No. 254.

Nelson, T. H. (1998). What's On My Mind? *First Wearable Computer Conference*, Fairfax, VA, May 12–13. [Available at: http://www.xanadu.com.au/ ted/zigzag/xybrap.html]

Nelson, T. H. (1999). *Welcome to ZigZag*. [Available at: http://xanadu.com/zigzag/tutorial/ZZwelcome.html

Nelson, T. H. (2001). Deeper Cosmology, Deeper Documents. *12th ACM Conference on Hypertext and Hypermedia*, Århus, Denmark, August 14–18.

Nelson, T. H. (2004). A Cosmology for a Different Computer Universe: Data Model, Mechanisms, Virtual Machine and Visualization Infrastructure. *Journal of Digital Information,* 5(1). Article No. 298, 2004-07-16. [Available at: http://jodi.ecs.soton.ac.uk/Articles/v05/i01/Nelson/]

Pangaro, P. (1997). *THOUGHTSTICKER: An Idiosyncratic History of Conversation Theory in Software, and its Progenitor, Gordon Pask*. Pangaro Inc., Belmont (MA). [Available at: http://www.pangaro.com/published/thstr-fest.html]

Pask, G. (1975). *Conversation, Cognition, and Learning*. Elsevier, New York.

Pask, G. (1976a). Conversational Techniques in the Study and Practice of Education. *British Journal of Educational Psychology,* 46, 12–25.

Pask, G. (1976b). *Conversation Theory: Applications in Education and Epistemology.* Elsevier, Amsterdam, Oxford.

Pask, G. (1988). Learning Strategies, Teaching Strategies, and Conceptual or Learning Style. In: R. R. Schmeck (Ed.), *Learning Strategies and Learning Styles* (pp. 83–99). Plenum Press, New York .

Pask, G. (1990). Some Formal Aspects of Conversation Theory and Lp. In: F. Heylighen, E. Rosseel and F. Demeyere (Eds.), *Self-Steering and Cognition in Complex Systems: Toward a New Cybernetics* (pp. 240–247). Gordon and Breach, New York .

Pask, G. and Gregory, D. (1986). Conversational Systems. In: R.J. Zeidner (Ed.), *Human Productivity Enhancement*, Vol. II (pp. 204–235). Praeger, New York.

Raina, M. (1979). Education of the Left and Right. *International Review of Education,* 25, 7–20.

Samples, B. (1977). Mind Cycles and Learning. *Phi Delta Kappan*, May.

Scott, B. (2001). Conversation Theory: A Constructivist Dialogical Approach to Educational Technology. *Cybernetics and Human Knowing,* 8(4), 1–25.

Spink, A. and Greisdorf, H. (1997). Users' Partial Relevance Judgments During Online Searching. *Online and CDROM Review,* 21(5), 271–280.

Strauss, A. and Corbin, J. (Eds.) (1997). *Grounded Theory in Practice*. London: Sage.

Zimmer, R. S. (2001). Variations on a String Bag. Using Pask's Principles for Practical Course Design. *Kybernetes,* 30(7/8), 1006–1023.

SECTION III

CIR PROCESSES

CHAPTER 6

A MULTITASKING FRAMEWORK FOR COGNITIVE INFORMATION RETRIEVAL

AMANDA SPINK
Queensland University of Technology

CHARLES COLE
McGill University

1. INTRODUCTION

Information retrieval (IR) research has developed considerably since the 1950's to include consideration of more cognitive, interactive and iterative processes during the interaction between humans and IR or Web systems (Ingwersen, 1992, 1996). Interactive search sessions by humans with IR systems have been depicted as interactive IR models (Saracevic, 1997). Human-IR system interaction is also modeled as taking place within the context of broader human information behavior (HIB) processes (Spink et al., 2002).

Research into the human or cognitive (user modeling) aspects of IR is a growing body of research on user interactivity, task performance and measures for observing user interactivity. The task context and situational characteristics of users' searches and evaluation have also been identified as key elements in a user's interaction with an IR system (Cool and Spink, 2002; Vakkari, 2003).

Major theorized interactive IR models have been proposed relating to the single search episode, including Ingwersen's (1992,1996) Cognitive Model of IR Interaction, Belkin et al.'s (1995) Episodic Interaction Model, and Saracevic's (1996,1997) Stratified Model of IR Interaction. In this chapter we examine Saracevic's Stratified Model of IR Interaction and extend the model within the framework of cognitive IR (CIR) to depict CIR as a multitasking process. This chapter provides a new direction for CIR research by conceptualizing IR with a multitasking context.

The next section of the chapter defines the concept of multitasking in the cognitive sciences and Section 3 discusses the emerging understanding of multitasking information behavior. In Section 4, cognitive IR is depicted within a multitasking framework using Saracevic's (1996, 1997) Stratified Model of IR Interaction. In Section 5, we link information searching and seeking models together, via Saracevic's Stratified Model of IR Interaction, but starting with a unitask model of HIB. We begin to model multitasking in cognitive IR in Section 6. In Sections 7 and 8, we increase the complexity of our developing multitasking model of cognitive IR by adding coordinating mechanisms, including feedback loops. Finally, in Section 9, we conclude the chapter and indicate future directions for further research.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 99–112.

2. CONCEPT OF MULTITASKING

Multitasking is the ability of humans to simultaneously handle the demands of multiple tasks through task switching (Just et al., 2000; Rubenstein, Meyer, and Evans, 2001). Cognitive psychologists have for decades studied many aspects of multitasking or task switching (Miyata and Norman, 1986; Carlson and Sohn, 2000). Although multitasking continues to be an important research area for technologies designers in general, many interactive technologies do not provide effective support for managing multitasking behaviors (Wickens, 1992). However, due to the increasing complexity of the global information environment people are increasingly engaged in multitasking and information task switching behaviors.

There are positive and negative aspects of multitasking. Rubinstein, Meyer and Evans (2001) found that multitasking between different types of tasks can reduce productivity. Wickens (1992), on the other hand, suggests that *time sharing* allows the simultaneous performance of multiple tasks and time swapping allows the sequential performance of tasks. Our position is that multitasking is a critical human behavior that allows people to cope with ever more complex environments by handling multiple tasks through task switching (Burgess, 2000; Carlson and Sohn, 2000; Lee and Taatgen, 2002).

3. MULTITASKING INFORMATION BEHAVIOR

When information is added to the mix, the concept of multitasking takes on an added layer of complexity. Multitasking information behavior may involve a combination of cognitive and physical actions, on dual or multiple tasks concurrently or sequentially, including switching between different information tasks. Cognitively, humans may sequence their thinking on information tasks and information task switching at different levels of complexity and speed. CIR is embedded within multitasking information behaviors that occur when users juggle the challenge of searching on multiple topics during the same search session.

IR systems users may engage in multitasking information behavior in two ways. First, a user may begin their IR system interaction with multiple topics, or second begin with a single topic and then develop additional topics during the search process. Both processes include information task switching, or switching back and forth between different topics during a search session. For example, a user may switch between seeking health information and new car information as they think and work on multiple information problems concurrently.

Our view, however, is that information searching, especially on the Internet, involves a much wider series of behaviors than purposive information seeking—thus we feel the term HIB is more appropriate. An example of an HIB perspective on user information seeking and searching within a multitasking framework view of cognitive IR is “vanity searching,” where the user puts his or name into the IR system.

A recent poll shows that 27 percent of Internet users conduct this sort of searching behavior (AOL Europe/ RoperASW Cyberstudy on Community, 2003). It is a searching behavior, but also a wider, HIB, because it serves no task-specific searching purpose; rather vanity searching plays into the searcher’s sense of place in a sociological context or system. This perhaps has motivational purposes by making sense of the user's search for information for a group of tasks, and is an example of how the Affective level in Saracevic's Stratified Model interplays with both the Cognitive level on the user side and the Content level on the computer side of the Model.

Currently, IR systems necessitate that users search sequentially and are designed to largely support only limited types of searching by specifying queries using terms to select documents or Web sites to fulfill a single information task. However, people in fact accomplish searching in much more complex ways than just this method of query specification and selection.

A growing body of studies shows that people often have many information problems/topics/tasks concurrently. Recent studies suggest that users' searches may have multiple goals or topics and occur within the broader context of their information-seeking behaviors (Spink, Ozmutlu, and Ozmutlu, 2002; Spink, 2004). People may pool their topics together and interact with an IR system on more than one related or unrelated topic. Overall, a user's single session with an IR system consists of seeking information on single or multiple topics, and also switching among topics (Spink, Ozmutlu, and Ozmutlu, 2002). In 1999, Spink, Batemen, and Greisdorf (1999) found that eleven (3.8 percent) of the 287 Excite users responding to a Web-based survey reported multitasking searches.

Spink, Ozmutlu, and Ozmutlu (2002) show that IR searches often include multiple topics, during a single search session or *multitasking search*. They found that multitasking information seeking and searching is a common human behavior as many IR system users conduct information seeking and searching on related or unrelated topics. In addition, Web or IR multitasking search sessions are longer than single topic sessions with mean topics per Web search ranging from 1 to more than 10 topics and a mean of 2.11 topic changes per search session.

Recent studies have examined multitasking searching on the Excite and AlltheWeb.com Web search engines (Ozmutlu, Ozmutlu, and Spink, 2003a, b). Ozmutlu, Ozmutlu, and Spink (2003a) provide a detailed analysis of multitasking sessions on AlltheWeb.com. They found that almost one third of AlltheWeb.com users perform multitasking Web searching. Multitasking Web search sessions often included more than three topics per session, are longer in duration than regular searching sessions, and most of the topics in multitasking searches were switching among general information, computers and entertainment.

Ozmultu, Ozmutlu, and Spink (2003b) found that multitasking Web searches are a noticeable user behavior, as one tenth of Excite users and one third of AlltheWeb.com users conducted multitasking searches. Multitasking Web search sessions are longer than regular search sessions in terms of queries per session and duration, with both Excite and AlltheWeb.com users searching for about three topics per multitasking session and submitting about 4-5 queries per topic.

Spink and Park (in press) studied multitasking information and non-information behaviors by business consultants. Key findings include: (1) seeking information formed 10.5% of business consultant daily tasks, (2) information seeking tasks occurred within multitasking and task switching sequences with computing and communication tasks, and (3) information seeking tasks often occur to support or respond to communication or computing tasks. Spink and Park (2005) provide a model of multitasking and task switching during HIB that includes cognitive, cognitive style and individual differences variables.

Typical Web search sessions are two queries, and some Web search sessions contain three or more queries (Spink and Jansen, 2004). Spink, Park, and Jansen (In press) present findings from two studies of multitasking during Web search. First, a study of two-query search sessions on the AltaVista Web search engine, and second, a study of three or more query search sessions on the AltaVista Web search engine. They examined the degree of multitasking search and information task switching during these two sets of AltaVista Web search sessions. A sample of two-query and three or more query sessions were filtered from AltaVista transaction logs from 2002 and qualitatively analyzed. Sessions ranged in duration from less than a minute to a few hours. Findings include: (1) 81 percent of two-query sessions included multiple topics, (2) 91.3 percent of three or more query sessions included multiple topics, (3) there are a broad variety of topics in multitasking search sessions, and (4) three or more query sessions sometimes contained frequent topic changes.

4. MULTITASKING VIEW OF CIR

Saracevic's Stratified Model of IR interaction (1996, 1997) implies a multitasking interaction between user and system, in which users engage in multiple tasks in order to construct a search interaction. Although Saracevic's model describes multitasking in terms of the user "getting the system to work" for one search task only, this model provides an effective start point for modeling the broader notion of multitasking we wish to describe here.

Therefore, as a result of our growing understanding of multitasking information behavior, we extend Saracevic's Stratified Model of IR interaction (1996, 1997) which depicts IR interaction as a dialogue between the participants in user-IR system interaction, the *user* and *computer* (system), through an interface at a *surface* level. Each of the participants is depicted as having different levels or strata. IR interaction is depicted as the interplay between various levels.

On the user side elements involve at least these levels: cognitive, affective, and situational. On the systems side elements involve at least these levels: engineering, processing and content. The user and the system side are mediated through the surface level or interface. Saracevic (1997) states "we can think of interaction as a sequence of processes occurring in the several connected levels or strata" (p. 315).

By thinking of multitasking as information behavior on two levels, we begin extending Saracevic's Stratified Model from multitasking so that the user can "get the system to work" to multitasking so that the user can get the system to work *and* to engage in searching of multiple topics during one search session. By doing so, this chapter extends previous cognitive/interactive IR approaches and proposes that interactive IR is a contextual interplay of multitasking behavior at different levels.

First, on an interactive search task level, people construct an interactive IR session as a series of tasks, including an embedded interplay of information problem, interactive search and other tasks. For example, embedded between telephoning and computing tasks, a search engine user coordinates many tasks when looking for medical information, such as translating their information problem into a set of search terms and strategy, search engine and search term selection, relevance judgments, etc.

On a second level, people engage in multitasking information behaviors or are seeking information on more than one topic concurrently. For example, a search engine user switches between seeking fashion information and medical information. Interactive IR can be conceptualized as interplay between different types of tasks and often different information problems. Conceptualizing interactive IR as a multitasking process within different levels embeds interactive IR within the broader framework of multitasking research in the cognitive/behavioral sciences. Effective interactive IR is successfully coordinating switching between related tasks. Web search can also include information multitasking behaviors that occur when users juggle the challenge of searching on multiple topics.

Multitasking information behavior research is a significant area of study for future research. However, in the IR context, multitasking behavior is still largely under-researched. Task analysis in cognitive IR is a new area of research. However, despite the new focus on tasks (Vakkari, 2003), current models of interactive IR do not consider multitasking behaviors.

Let us take but one part of user multitasking, the user "getting the system to work." Current search technologies are designed to largely support only limited types of searching by specifying queries using terms to select documents or Web sites to fulfill a single information task. However, people in fact accomplish interactive IR in much more complex ways than just this method of query specification and selection. In fact, term selection and query formation, even to search the IR system for one single task or unitask searching, takes an enormous amount of cognitive coordination activity. Therefore, understanding and modeling multitasking information behaviors requires us to start with an analysis of the unitask search, to acquire an understanding of the coordination and interplay between information problem, interactive search and other tasks.

People must coordinate the translation of their information problem(s) by performing search term selection tasks, tactic and strategy tasks, search engine interaction tasks, relevance judgments, etc. Research shows that humans have different levels of cognitive coordination (Miyata and Norman, 1986). The Saracevic Stratified Model also depicts different levels of cognitive coordination. Interactive IR occurs as a series of coordinated task actions. To achieve interactive IR, humans *coordinate* a number of tasks, including their cognitive state, level of domain knowledge, and their understanding of their information problem, into a coherent series of activities that may include searching, seeking, interactive browsing and retrieving and constructing information.

Humans cognitively coordinate their information seeking level behaviors with their interactive searching level (human-system interaction) behaviors; including the recognition and making sense of and cognitively articulating an information problem or a gap in their knowledge. Humans then coordinate these processes to construct an interactive IR process embedded within their broader information seeking, HIB, and non-information behaviors.

Establishing then sustaining an effective interactive IR process requires humans to coherently coordinate and multitask their information problem and interactive search tasks. In other words, an information seeker must coordinate a number of tasks, including their cognitive state, level of knowledge, their understanding of their information problem, into a coherent series of sustained activities that include seeking, searching, retrieving and using information.

We know that hand-eye coordination is a physiological process that humans develop from childhood. Research appreciates that hand-eye coordination as well as physiological processes, also involves cognitive processes, and that there is an overriding coordinating process directing these complex processes or behaviors. Nevertheless, we have very little understanding of how humans learn the process of cognitively coordinating their information problems into coherent processes of HIB and interactive IR.

5. STRATIFIED MODEL OF INFORMATION SEARCHING AND INFORMATION SEEKING

To understand coordination, we use Saracevic's Stratified Model of IR interaction—a searching model—to analyze the multitasking searching of the user, which requires coordination between problem situation, query formulation and search tactics, etc, inside the larger multitasking framework of multiple topic information seeking during a single search session. We wish to remain with Saracevic's Stratified Model, using it to conceptualize a linkage between information searching and seeking. We do this via the other two levels of the user side of Saracevic's interaction model: the affective and situation levels. These two levels, with the cognitive level already referred to, complete the three levels of the user side of Saracevic's Stratified Model.

The traditional interest of library and information science (LIS) research has been the contemporary human activity of actively seeking information, such as people coming to the library to seek information. But studies since at least Line (1971) have shown that non-library methods of user information seeking, where the user seeks information through so-called informal channels like talking to relatives or work colleagues, were preferred by the user over accessing information systematically through formal channels made available to the user in the library. As a result, LIS research has reconfigured its focus over the last twenty to thirty years, contextualizing its perspective on information behavior outward from the human's face-value or expressed information need to starting from the user's "problem situation" (Belkin, 1980; Ingwersen, 1996; Case, 2002)—i.e., information need starts from recognition of a problem situation (Wersig, 1971).

As a result, the current dominant ethos in information seeking research is the study of purposive behavior within a problem solving framework borrowed from the cognitive science approach to models of general problem solving (Case, 2002; Marchionini, 1995). The problem solving approach presents a specific view of information-seeking as goal-directed behavior, with the resolution of the problem and/or the presentation of the solution as that goal. The problem solving approach conceptualizes the process of information seeking as the human behavior that supports the process of information construction, which is accomplished in stages. Recent information seeking studies support the notion that users seek information in a series of stages that link to the performance of a task/project, and that they adopt different strategies and exhibit different information behaviors at different stages of their information-seeking process (Kuhlthau, 1993).

A classic example of stage approach to problem solution is Popper (1975) in his problem solving schema of conjectures and refutations. The problem solver starts off conscious information seeking behavior in an initial state of understanding of the problem, called P_1. The problem solver, after interacting with information objects, arrives at a tentative theory (TT), corrects it for errors (EE) by engaging in further interaction with information objects, then finally arrives at a second state of understanding the problem, called P_2.

To quantitatively chart the user's course through the problem stages, Wilson et al. (2002) adapt the LIS concept of uncertainty, specifically Kuhlthau's (1993) concept of uncertainty, to test whether a person's uncertainty decreases from the beginning to end of the problem solving process (Pennanen and Vakkari, 2003).

Kuhlthau's (1993) concept of uncertainty is multidimensional, having an affective as well as a cognitive part (Cole, 1997). Cognitive uncertainty is created in the information seeker when an outside message (e.g., from an IR system) with a requisite proportion of uniqueness" (Kuhlthau, 1993, p. 30) causes in the seeker cognitive uncertainty, defined as "a lack of understanding, a gap" (Kuhlthau, 1993, p. 11); the cognitive uncertainty, in turn, causes affective uncertainty (i.e., doubt) which initiates the information seeking behavior in the individual.

The problem solving model of Popper (1975) addresses the situation level of Saracevic's Stratified Model of interactive IR, while the information seeking model of Kuhlthau (1993) addresses the affective (as well as cognitive) level of the Saracevic model. As well, both the Popper and Kuhlthau models describe and explain a stage model of task completion.

For our present purpose of developing a multitasking and coordinating approach for CIR, we start by defining the level of information seeking and the level of information search together as being the object of the key sustaining process of CIR. Information seeking, which is the traditional sphere of study and research in LIS, is purposive HIB. Information searching is a specific activity within information seeking, a narrow definition of which is user-IR system interaction HIB (e.g., query formulation, user strategies and tactics, problems while searching for information using an IR system, etc.).

Humans cognitively coordinate their information seeking level behaviors with their information searching level behaviors; including the processes of the recognition and making sense of, and cognitively articulating an information need or a gap in their knowledge. Humans coordinate these processes to construct an information-seeking process, part of which is to go to an IR system to conduct an information search. Coordination is also related to movement through a human information seeking process towards the performance of a goal or task or the resolution of a problem, as shown in Figure 1.

In Figure 1, we use Kuhlthau's (1993) six stage ISP model and Popper's (1975) schema of problem resolution as models of task performance and problem solution respectively, to begin to diagram human information coordinating behavior (HICB).

Figure 1 is a very linear conception of the information seeking and information searching process where HIB is reduced to a production line in a factory to show stages of a unitask performance only. But even in this unitask conceptualization of these processes, humans coordinate various elements in order to move through their information-seeking behavior process. An information search of an interactive IR system involves Saracevic's three strata or levels on the user's side of the interaction: the cognitive, affective and situational.

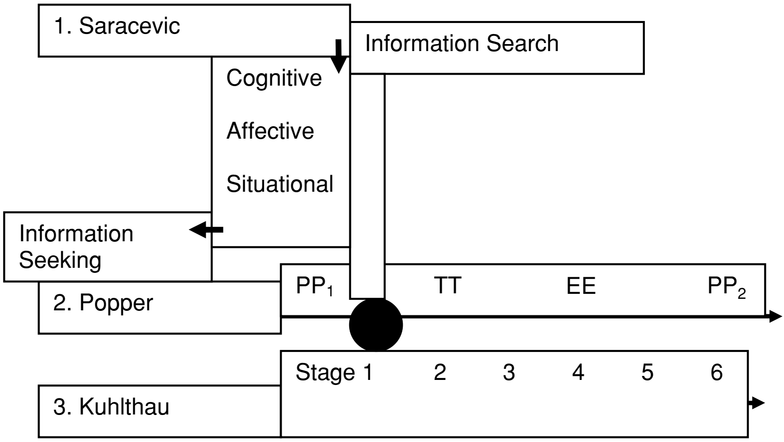

Figure 1: Information seeking & search sustained by HICB (small filled-in circle in middle of diagram)—forward progress toward information task completion

In Figure 1, these three levels link the search model of Saracevic and the seeking models of Popper and Kuhlthau. Part of the information-seeking process is the users' translation of their information problem into a form, such as a query to an IR system, that allows them to construct information from texts, documents and Web sites. Translating the elements of their information problem to effective information-seeking and searching strategies and tactics is essential to an effective coordination process.

For example, a human is seeking information on a particular disease. They enter a library or begin to search the Web. To enable their information-seeking and searching process to move forward they must understand the dimensions of their information problem and coordinate their information seeking and searching processes to the point where they are able to interact with the functional structure of the library or Web system. The coordination process between information problem and the information-seeking and searching processes must take place before a human enters a keyword into the Web or begins to browse the library shelves.

Establishing and sustaining an effective information seeking and searching process require humans to develop effective and coherent information coordinating behaviors and processes. In other words, information seekers must coordinate a number of elements, including their cognitive state, level of knowledge, and their understanding of their information problem, into a coherent series of sustained activities that may include seeking, searching, retrieving and using information. We know that hand-eye coordination is a physiological process that humans develop from childhood. But how do humans learn the process of coordinating their information needs into coherent processes of human information seeking, searching, retrieving and use behaviors?

6. MULTITASKING CONCEPTION OF HIB

Let us now add the additional element of multitasking for multi-topics during one search session: seeking information for multiple topics or multiple information needs. That is, in addition to coordinating HIB processes for one task, users actually multitask while engaging in information seeking and searching. For example, an information seeking process switches between seeking fashion information and medical information. Information seeking can be conceptualized as the interplay between different types of tasks and often different information problems.

Rather than seeing this as a negative or counterproductive, as does recent research in driving while engaging in another task (Just, et al., 2000; Rubenstein, Meyer, and Evans, 2001), we see multitasking and the information seeking/searching processes that allow multitasking to occur in a broader, more holistic perspective. We have referred to this holistic perspective earlier as human information behavior (HIB) (Spink and Cole, in press). HIB conceptualizes multitasking as a necessary part of the success of information seeking/searching processes. However, in the information behavior context, multitasking information behavior is still largely under-researched. In addition, current HIB models are based in a single information task paradigm. But, information behaviors are accomplished by people in much more complex ways.

7. COGNITIVE COORDINATION MECHANISMS

In this section we look at Saracevic's (1996, 1997) Stratified Model of Interactive in terms of coordinating multitasking searching and seeking. Saracevic's Model has three levels or strata on the user side and three levels or strata on the computer side, which provides a multidimensional portrait of the actors and levels of user-system interaction.

On the user side of Saracevic's Model, the cognitive, affective and situational levels are the object of our analysis for incorporating multitasking coordinating activities, including feedback loops, into cognitive IR system design. A specific kind of coordinating activity, feedback loops, are described in Section 8 below. On the computer side of Saracevic's Stratified Model, we focus on the content level, which includes information sources contained in the IR system database: "users interact with the 'texts' (including images, data and their representations) in the information resources" (Saracevic and Spink, 1997, p. 744).

This interplay between the cognitive, affective and situational levels on the user side, and the content level on the computer side of Saracevic's Stratified Model, indicates not only the need for a coordinating mechanism, but also assumes that the utilization of the levels is part of the coordinating mechanism invoked when a user seeks and searches for information while using an IR system. That is, the user, perhaps unconsciously, accesses the other levels to facilitate the coordination of information processing, task switching and other HICBs during multitasking (Spink, Park, and Cole, 2005; Cole and Leide, 2005).

We provide a hypothetical example in support of this last statement. In Section 3 (above), we referred to "vanity searching." Vanity searching is a HIB in the context of multitasking and cognitive IR because it serves no obvious purpose in performing a task or solving a problem. However, it does facilitate coordination of multitasking by allowing the user to access his or her affective and situational levels of searching and seeking. Accessing these levels in-between purposive searching and seeking may, we conjecture in Section 3, have motivational or "making sense" attributes for the individual doing the searching and seeking. In this way, accessing the user's affective and situational levels may be a coordinating mechanism for multitasking.

Allowing the individual to access affective and situational levels of searching and seeking, points out the mundane necessity for a switching mechanism as part of the coordination. The physiological process of hand-eye coordination develops in humans from childhood, and these physiological processes are related to many and well-studied multitasking activities of humans. But how do humans learn the process of cognitively coordinating their varied information problems of work and everyday life into coherent HIB processes during interactive information seeking and interactive information searching?

To overcome coordination problems, information seekers must perform additional work in the form of *cognitive coordination mechanisms*, which includes task switching as shown in Figure 2. Coordination mechanisms may be quite specific, such as making a relevance judgment about potential information. Relevance judgment making is a switching mechanism from information search to information construction (or lack thereof), and either back to information searching or to other multitasking activities.

As well, cognitive coordination mechanisms may be quite general, such as hierarchical mechanisms to manage assignment of activities, including prioritization. Note that many cognitive coordination mechanisms are primarily information processing activities (e.g., the selecting/evaluation of information measured against some ideal of what the user wants stored in memory, ordering or picking tasks, negotiating with other actors or informing them about planned activities) and are thus potential candidates for support from electronic media.

In general, there may be several cognitive coordination mechanisms that could be used to address a given dependence—i.e., different information seekers may use different mechanisms to address similar problems, thus resulting in a different information behavior process. Given a particular information seeker performing an information task, one way to generate alternative processes is to identify the particular dependences and cognitive coordination problems faced by the information seeker and consider what alternative cognitive coordination mechanisms could be used to manage them.

8. FEEDBACK LOOP

A cognitive coordinating framework for HIB during multitasking is a complex human process to model. The exploration of coordination is crucial to the development of a science of information, a theory of HIB. An important theoretical foundation for a human information coordinating behavior (HICB) process is information feedback (Spink and Saracevic, 1998). The feedback loop allows humans to evaluate the effect of one information behavior on another, which includes the effect of an HIB on achieving the goals and objectives of the overall HIB activity. With the feedback loop, humans come to know whether they should continue on with or modify the HIB, select an alternate HIB, or leave the task for later.

Part of the feedback importance, we believe, is multitasking because the split second decision to leave a task automatically turns one task and one information seeking/searching process, into a multitask process. However, in information seeking and searching, leaving a first task to go to a second may enrich the processes involved in

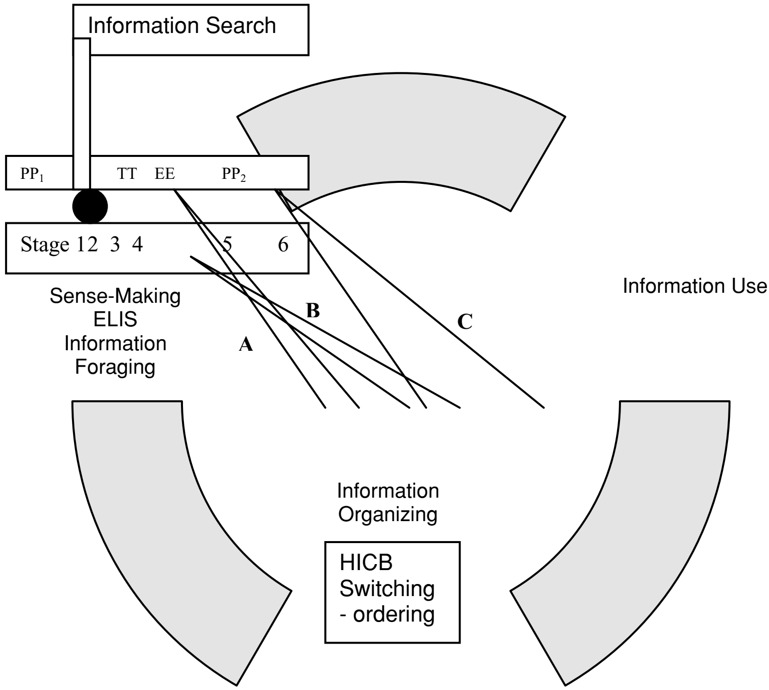

Figure 2: Multitasking and coordinating model for HIB including CIR

completing the first task. During information seeking and search, multitasking, rather than being discouraged, as perhaps in driving, should be encouraged. The information feedback loop forms the basis of the looping HICB process. But how do humans achieve through feedback loops the necessary and sustaining coordination, including recognizing, making sense of and cognitively articulating an information need or gap in their knowledge, while at the same time constructing and maintaining an information seeking-searching process that leads to a finite end? The informational feedback loop of coordination behavior is one of the most penetrating fundamentals in the information sciences and must form part of the basis for any theory of HIB.

Figure 2 incorporates various elements into an integrated model of CIR that includes task switching.

Figure 2 primarily consists of coordinating the switching between tasks. We also incorporate the Figure 1 information seeking and searching diagram in Figure 2, substituting the diagram for the words "Information seeking" and "searching" in Figure 1. We simply overlay the diagrams of the switching behavior over the information seeking and searching model. The triangles in the diagram indicate feedback loops, from one task, to the cognitive coordination behavior of switching, to a second task. Given the complexities of creating a multitasking HIB model, in Figure 2 we have made a first attempt to develop a multitasking and coordinating model for HIB.

9. CONCLUSION AND FURTHER RESEARCH

In this chapter, we have presented a new, complex perspective on cognitive IR, expanding the traditional IR perspective to include HIB's that are broader than the information search and seeking behaviors usually associated with IR. As we have described them here, these HIBs serve as coordinating behaviors allowing IR users to engage in complex multitasking searching and seeking activities. Rather than seeing multitasking as a negative feature of user performance while accessing information using IR systems, we view multitasking as the natural way of searching and seeking information. IR systems that include these cognitive, affective and situational elements of multitasking searching and seeking in their design should, we believe, enhance user performance in a variety of ways.

This chapter has proposed that, theoretically and practically, interactive IR can be conceptualized as a multitasking and coordinating process at various levels as the interplay of information problem and interactive search tasks. Exploring multitasking and coordination behaviors designate relatively new and heuristic directions for interactive IR research. The authors are currently conducting further studies to extend our understanding of the nature, patterns and impacts of interactive IR as multitasking. Currently, Web systems and interfaces provide limited support for multitasking search. As the complexity of information structures and problems increases, more complex human information processes and more effective Web technologies are required to sustain effective HIBs.

Further research is being conducted to investigate:

1. How multitasking searches differ from non-multitasking sessions.
2. If a user's information-seeking stage affects the number and performance of multitasking searches.
3. The relationship between the nature of the user's information problems and multitasking, and how they are coordinated.
4. The interplay between information and non-information tasks (Spink and Park, forthcoming).

REFERENCES

AOL Europe/ Roper ASW Cyberstudy on Community (2003). Last visited April 22, 2005: http://www.timewarner.com/corp/newsroom/pr/0,20812,669749,00.html.

Belkin, N. J. (1980). Anomalous States of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science,* 5, 133–134.

Belkin, N. J., Cool, C., Stein, A., and Theil, U. (1995). Cases, Scripts, and Information Seeking Strategies: On the Design of Interactive Information Retrieval Systems. *Expert Systems with Applications,* 9(3), 379–395.

Burgess, P. W. (2000). Real-World Multitasking From a Cognitive Neuroscience Perspective. In: S. Monsell and J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII*. The MIT Press, Cambridge, MA.

Case, D. O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior*. Amsterdam: Academic Press.

Carlson, R. A. and Sohn, M.-Y. (2003). Cognitive Control of Multistep Routines: Information Processing and Conscious Intentions. In: S. Monsell and J. Driver (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII.* The MIT Press: Cambridge, MA.

Cole, C. (1997). Information as Process: The Difference Between Corroborating Evidence and "Information" in Humanistic Research Domains. *Information Processing and Management,* 33, 55–67.

Cole, C. and Leide, J. E. (2005). A Cognitive Framework for Human Information Behavior: The Place of Metaphor in Human Information Organizing Behavior. *New Directions in Human Information Behavior*. Dordrecht: Springer.

Cool, C. and Spink, A. (2002). Information Retrieval (IR) in Context. *Information Processing and Management*, 38(5), 605–611.

Ingwersen, P. (1992). Information Retrieval Interaction. London: Taylor Graham.

Ingwersen, P. (1996). Cognitive Perspective of Information Retrieval Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation,* 52(11), 3–50.

Just, M. A., Carpenter P. A., Keller, T. A., Emery, L., Zajac, H., and Thulborn, K. R. (2000). Interdependence of Non-Overlapping Cortical Systems in Dual Cognitive Tasks. *Neuroimage,* 14, 417–426.

Kuhlthau, C. C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex.

Lee, F. J. and Taatgen, N. A. (2002). Multitasking as Skill Acquisition. In *Proceedings of CogSci 2002: Annual Meeting of the Cognitive Science Society, August 2002, Fairfax, VA*.

Line, M. B. (1971). The Information Uses and Needs of Social Scientists: An Overview of INFROSS. *Aslib Proceedings*, 23 (August), 412–435.

Marchionini, G. (1995). *Information Seeking in Electronic Environments.* New York: Cambridge University Press.

Miyata, Y. and Norman, D. (1986). Psychological Issues in Support of Multiple Activities. In: D. A Norman and S. W. Draper (Eds.), *User Centered Design*. Lawrence Erlbaum, NJ.

Ozmutlu, S. Ozmutlu, H. C. and Spink, A. (2003a). Multitasking Web Searching: Implications for Design. *ASIST'03: Annual Meeting of the American Society for Information Science and Technology, Oct 18–22, 2003. Long Beach, CA*.

Ozmutlu, S., Ozmutlu, H. C., and Spink, A. (2003b). A Study of Multitasking Web Searching. *IEEE ITCC'03: International Conference on Information Technology: Coding and Computing, April 2003, Las Vegas, NV.*

Pennanen, M., and Vakkari, P. (2003). Students' Conceptual Structure, Search Process, and Outcome While Preparing a Research Proposal: A Longitudinal Case Study. *Journal of the American Society for Information Science and Technology,* 54(8), 759–770.

Popper, K. R. (1975). *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

Rubinstein, J., Meyer, D., and Evans, J. (2001). Executive Control of Cognitive Processes in Task Switching. *Journal of Experimental Psychology,* 27(4), 763–797.

Saracevic, T. (1996). Modeling Interaction in Information Retrieval (IR): A Review and Proposal. In *Proceedings of the Annual Meeting of the American Society for Information Science,* 33 (pp. 3–9). Baltimore MD.

Saracevic, T. (1997). The Stratified Model of Information Retrieval Interaction: Extension and Applications. In *Proceedings of the Annual Meeting of the American Society for Information Science,* 34 (pp. 313–327). Washington DC.

Saracevic, T. and Spink, A. (1997). Interaction in Information Retrieval: Selection and Effectiveness of Search Terms. *Journal of the American Society for Information Science*, 48(8), 741–761.

Spink, A. (2004). Multitasking Information Behavior and Information Task Switching: An Exploratory Study. *Journal of Documentation,* 60(4), 336–345.

Spink, A., Bateman, J., and Greisdorf, H. (1999). Successive Searching Behavior During Information Seeking: An Exploratory Study. *Journal of Information Science,* 25(6), 439–449.

Spink, A. and Cole, C. (in press). Human Information Behavior: Integrating Diverse Approaches and Information Use. *Journal of the American Society for Information Science and Technology.*

Spink, A. and Jansen, B. J. (2004). *Web Search: Public Searching of the Web*. Kluwer Academic Publishers.

Spink, A., Ozmutlu, H. C. and Ozmutlu, S. (2002). Multitasking Information Seeking and Searching Processes. *Journal of the American Society for Information Science and Technology,* 53(8), 639–652

Spink, A. and Park, M. (2005). Information and Non-Information Task Interplay. *Journal of Documentation,* 61(4).

Spink, A. and Park, M. (in press). Multitasking Interplay of Information and Non- Information Tasks. *Journal of Documentation*.

Spink, A., Park, M. and Cole, C. (2005). Multitasking and Coordinating Framework for Human Information Behavior. In A. Spink and C. Cole (Eds.) *New Directions in Human Information Behavior*. Dordrecht: Springer.

Spink, A., Park, M. and Jansen, B. J. (in press). Multitasking During Web Search Sessions. *Information Processing and Management.*

Spink, A. and Saracevic, T. (1998). Human Computer Interaction in Information Retrieval: Nature and Manifestations of Feedback. Interacting with Computers: The International Journal of Human Computer Interaction, 10(3), 249–267.

Spink, A., Wilson, T. D., Ford, N., Foster, A. and Ellis, D. (2002). Information-seeking and Mediated Searching. Part 1. Theoretical Framework and Research Design. *Journal of the American Society for Information Science and Technology*, 53(9), 695–703.

Vakkari, P. (2003). Task-Based Information Searching. *Annual Review of Information Science and Technology*, 37 (pp. 413–464). Information Today: Medford, N.J .

Wersig, G. (1971). *Information-Kommuikation-Dokumentation*. Pullach bei Munchen: Verlag Documentation.

Wickens, C. D. (1992). *Engineering Psychology and Human Performance*. HarperCollins: New York.

Wilson, T. D., Ford, N., Foster, D., Ellis, D., and Spink. (2002). Information Seeking and Mediated Searching. Part 2: Uncertainty. *Journal of the American Society for Information Science and Technology,* 53(9), 704–715.

CHAPTER 7

EXPLANATION IN INFORMATION SEEKING AND RETRIEVAL

PERTTI VAKKARI AND KALERVO JÄRVELIN
University of Tampere, Finland

1. INTRODUCTION

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

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

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

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

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 113–138.

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

The goals of IS&R research may be classified as (a) theoretically understanding information seeking and retrieval in the form of models and theories, (b) empirically describing and explaining IS&R in various contexts, and (c) providing support in the design of information systems and information management in various contexts. The whole area is rather pragmatic: many experimental (IR) studies aim at improving IR system effectiveness; many other studies are descriptive regarding the ways people access information (but these studies also have improving people's information access on their agendas).

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

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

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

2. AN APPROACH TO GROWTH OF KNOWLEDGE

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

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

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

Substantial theories deal directly with the concrete research object, e.g. information search process or indexing. They essentially contain a set of concepts and assertions relating the concepts in an account of some phenomenon in IR. They include hypotheses that can be tested and are intended as answers to particular research questions (e.g., do weighted keywords retrieve more effectively than un-weighted ones?) The extent to which empirical testing provides support for the substantial theories determines the degree to which one may say that theoretical growth has occurred (Wagner and Berger, 1985).

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

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

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

- *Dependent* variables—the variation of which is being explained.
- *Independent* variables—the ones systematically varied in order to see the responses in the dependent ones.
- *Controlled* variables—the ones fixed to prevent uncontrolled variation in the results. They are not as central in describing the research object as the two former ones.

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

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

3. CURRENT IR RESEARCH AND ENRICHING TRADITIONAL APPROACHES TO IS&R

3.1. The Lab IR Framework

The basic laboratory IR framework has no user involvement. This model suggests documents, search requests, their representation, the database, queries, and the matching of the two latter as foci of research and development. Methodologically, it also suggests relevance assessments, recall base construction and query result evaluation as foci of analysis (Figure 1)

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

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

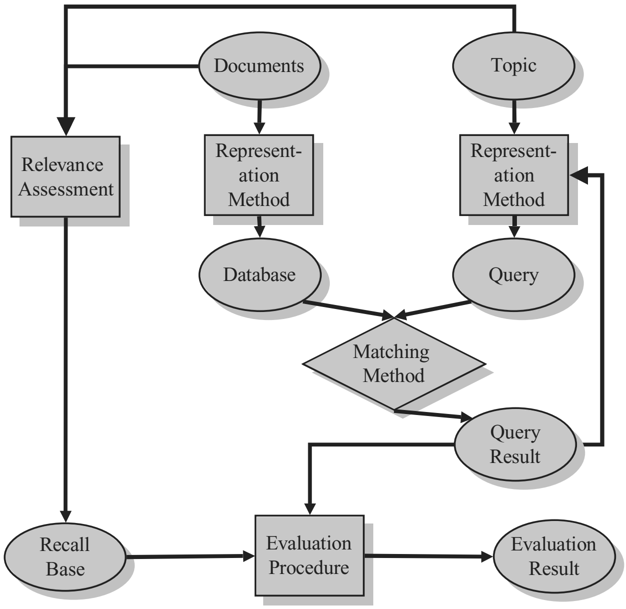

Figure 1: The laboratory framework of IR schematized

In the model (Figure 1) the major independent factors (documents, topics and their representations) and also the central component of the dependent factor (relevance assessments) are taken as given. They are the sole source of variation of precision and recall. The variance of explaining factors is the natural point of departure for explaining the variance of precision and recall. However, it can be presumed that factors outside this model influence the variation of the independent factors, and consequently of the dependent factors. If it is possible to identify factors which cause systematic variation in independent factors and also in dependent factors, it is necessary to enrich the model by these external factors, particularly if they help in enhancing the goal of IR evaluation, to improve retrieval effectiveness. This implies changes in the meta-theoretical assumptions of the model. In the following we explore the possible external factors to the model, which may cause systematic variation in both independent and dependent variables.

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

3.1.1. Dependent variables

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

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

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

Some of the problems of using higher-order relevance in IR systems, or their design, may be discussed in terms of Figure 2. A real user, being thrown into a situation, may well be able to recognize a relevant document once presented (hence the exclamation mark). However, he may have difficulty in discussing the relevance criteria of the task and situation. Furthermore, he certainly has difficulty in expressing a request and formulating a query to the IR system, at least anything other than topical *as far as text is concerned* (except for bibliographic fields etc., if available, as discussed below), because current systems do not provide for anything else. The system designer probably never had the slightest idea of anything other than explicit topical indexing features, because there is no known pattern of situational indexing features that are explicit in text (the computer does not handle implicit features) and useful to users (Cooper, 1971). Therefore the available indexing features may not correlate to the situational relevance

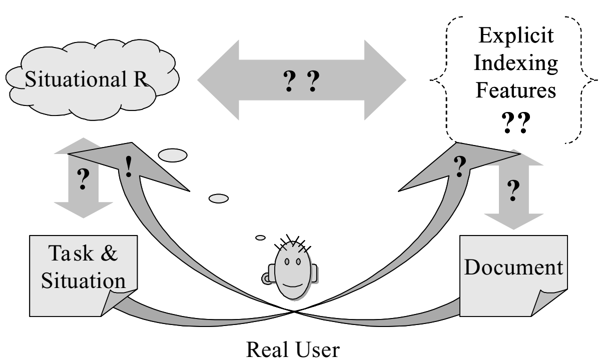

Figure 2: Situational R (relevance) in retrieval (Kekäläinen and Järvelin, 2002)

criteria, which the user did not express, save for one thing: topical relevance heavily correlates to situational as shown by Burgin (1992), and Vakkari and Hakala (2000).

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

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

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

3.1.2. Independent variables: Requests and their representations

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

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

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

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

3.1.3. Independent variables: Documents and their representations

A presupposition of the laboratory model seems to be that all textual documents are similar so that all can be represented for IR similarly. It is expected that, e.g., variation between document genres (like news or scientific articles) or domains (like engineering or art) in vocabulary or in word distribution in documents, does not matter in how documents should be indexed. It seems to us that specifying documents and their representations by differentiating between documents of various genres and domains is a fertile way of increasing the explanatory power of the model. This is an open question which can be answered by empirical studies. Evidently new IR techniques are needed, which react to the differences between genres.

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

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

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

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

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

3.2. Experimental Interactive IR

Belkin (1993) proposed that the central process in information retrieval is user interaction with text, and that the user is the central component of the IR system. He argued that in traditional IR research little emphasis has been laid on developing representation schemes specific to information needs and problems. He further argued that one of the main goals of IR systems is to support inquirers in their information search process, especially in their interaction with texts. He stressed that understanding and representing users' information needs and problematic situations is important in developing tools for supporting the search process.

As suggested in the previous chapters the traditional laboratory model could be enriched by introducing representations of typical task and problem situations from various domains. This enrichment of the laboratory model does not include human actors and excludes studying the search process and various retrieval tools for supporting information searching. This can be studied only in interactive IR (IIR). Figure 3 depicts the elements of the experimental IIR model.

3.2.1. Dependent variables

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

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

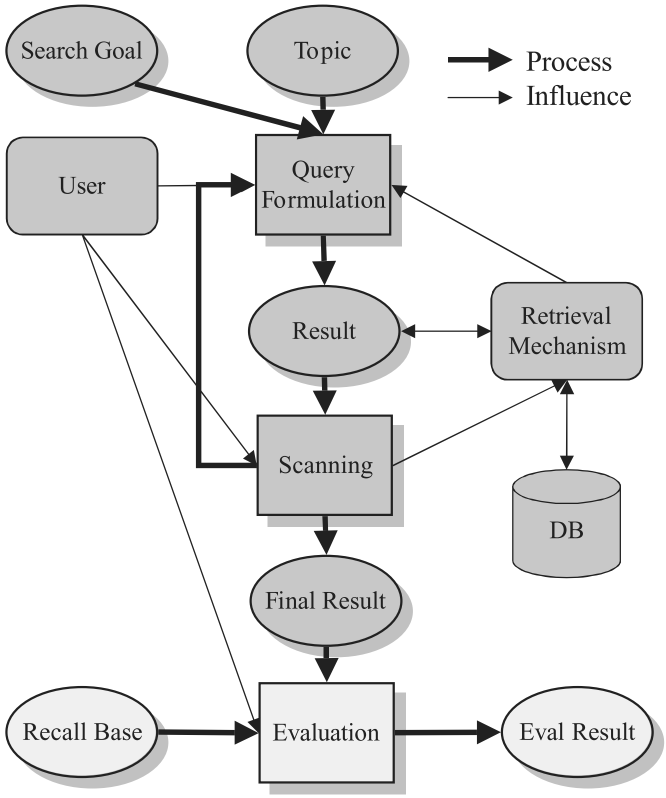

Figure 3: Schematized experimental IIR framework

3.2.2. Independent variables

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

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

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

Provision of term suggestion devices based either on search results or knowledge structures have been the most typical independent factor in IIR studies. Provision of semiautomatic or interactive term devices based on relevance feedback on the one hand (Fowkes and Beaulieu, 2000; Koenemann and Belkin, 1996), and thesauri or ontologies on the other hand (Jones et al., 1995; Suomela and Kekäläinen, 2005; Sihvonen and Vakkari, 2004) or both (Joho, Sanderson, and Beaulieu, 2004) have been studied. Also variation in query composition has been used as an independent factor. Supply of tools to support scanning and evaluating search results have been applied also to some extent as independent variables (Dumais, Cuttrell, and Chen, 2000).

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

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

3.3. Field Studies on IIR

The amount and range of field studies is so great that we give only a rough overview here. Field studies typically do not apply experimental research design. They use user generated search topics and goals. The research design includes typically dependent, independent and controlled variables. The range of factors used as dependent variables varies from search process variables to search outcome variables like the number of relevant items. The range of independent factors is still more comprehensive, varying from searchers' characteristics to search process variables. Controlled variables can be anything from searchers' characteristics to process variables. The one and the same variable may function in different roles in research designs depending on the research question. Borgman, Hirsh, and Hiller (1996) discuss these different variables.

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

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

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

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

3.4. Information Seeking Studies and IR

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

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

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

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

3.4.1. Kuhlthau's ISP model: An example of an information seeking study

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

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

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

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

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

Vakkari (2001) has extended Kuhlthau's ISP model in the field of interactive IR (Figure 4). The independent variable in both studies was stages in task performance (ISP). Some of the dependent variables, like information types and search tactics used by Kuhlthau, were modified and refined for IR purposes. For example, in the ISP model search tactics in information systems were categorized as browsing and querying, which is naturally too rough for observing in detail search tactics in databases. A more comprehensive and specific categorization of search tactics was developed. The ISP

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

Figure 4: Main concepts in Kuhlthau (1993) and Vakkari (2001)

model was enriched by introducing variables describing the use of search terms and operators, and relevance assessments. The basic hypothesis for field studies was that the stages of task performance were connected to the types of information searched for, to the changes of search terms and tactics, and to relevance judgements. These consequences from the ISP model were naturally based on the extension and refinement of its basic concepts for the purposes of IR.

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

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

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

4. DO WORK TASKS AFFECT IS&R?

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

4.1. A Possible Framework for Analysis

Basically, we approach IS&R as embedded contexts of retrieval, seeking and work tasks (Figure 5). IR serves the goals of seeking, and information seeking the goals of the work task (or other interest). The same person symbol in all three contexts denotes the same or another actor(s) performing the work task, the seeking task and the retrieval task, interpreting the tasks, performing the process and interpreting the outcome, possibly resulting in task reformulation in each context. The person symbol in the IR context signifies the possibility of applying human relevance feedback during a traditional two-run IR experiment as well as real longitudinal interactive IR over several short-term interactions.

Figure 5 admittedly focuses on IR and neglects other means of information access. Furthermore, it is not analytic about significant factors affecting the seeking and work processes (Pharo, 2004; Vakkari, 2003). However, possible evaluation criteria in each context are given at the arrows A–D. While many of them remain at quite an abstract level, they suggest types of variables, the variation of which one might want to explain. It allows for explaining variation in information access by organizational or work task criteria as well as explaining variation in work task performance by information access criteria. The following sections detail these possibilities.

4.2. Some Possible Study Designs

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

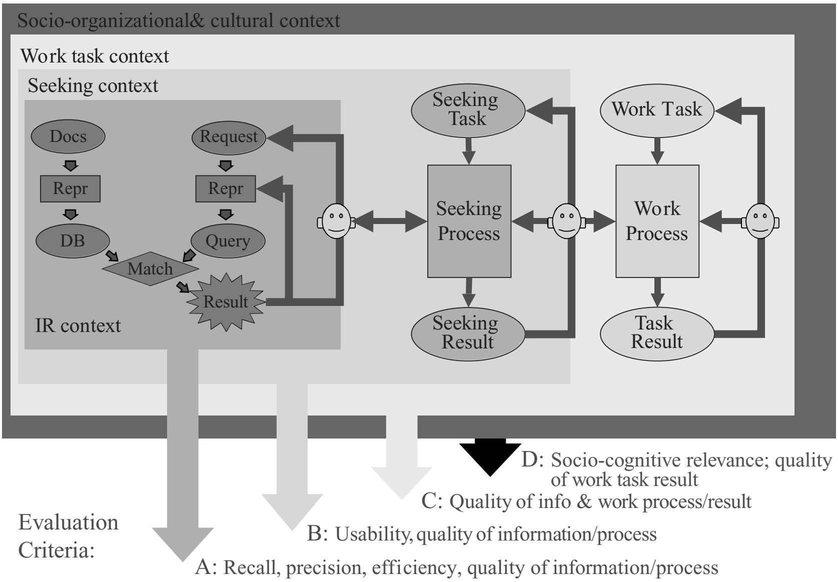

Figure 5: Nested contexts and evaluation criteria for task-based IS&R (extension of Kekäläinen and Järvelin, 2002; Järvelin and Ingwersen, 2004)

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

9. The *access and interaction dimension* covers strategies of information access, interaction between the actor and the interface (both in social and in system contexts).

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

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

4.2.1. Explaining the variation of recall and precision under novel conditions

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

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

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

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

4.2.2. Explaining the access process under work task conditions

This study design represents a departure from the explanation of the variation of recall and precision. Instead, it seeks to explain how real people solve their information problems related to different types of work tasks in realistic contexts. The underlying motivations for this effort include (1) the theoretical interest in better understanding the phenomenon of information access in work task contexts, and (2) the technical interest in developing novel and better kinds of task-based information access environments, which may require novel kinds of components. The study designs may be characterized by:

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

The studies may be operationalized under two approaches: field studies and laboratory studies. In the *field study* version data collection happens in a real working organization. However, the type of organizations may vary between studies, e.g., governmental domain (e.g., law, banking) and information environments (e.g., comprehensive in-house information systems available). Likewise, the actors may vary in their expertise. The *laboratory study* versions of this approach distil findings from the field studies for more rigorous empirical testing. This may involve a collection of standard work tasks as routinely used in business schools (Isenberg, 1986) and an artificial work environment in the lab.

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

4.2.3. Explaining the access process under actor conditions

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

- The independent variables: the actors performing the work tasks divided into groups in relation to work task performance (e.g., experienced professionals vs. novices regarding the work tasks, or one group receiving information access training plus a control group).
- The controlled variables: the organizational setting (e.g., a single large company), the actors performing the work tasks (e.g., all experienced professionals regarding the work tasks); the information environment (e.g., all the shared information access tools as above); a set of domain-specific work tasks controlled for their features—a representative mix or fixed type.
- The dependent variables: the choice to use each access tool; the application way of each access tool; the problems encountered; the kind of information sought with each tool; the kind of information identified and utilized; the quality of work task outcome.

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

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

4.3. Two Sample Studies

In this section, we discuss two studies that can be seen to utilize a task-based approach to IS&R. Neither study is based on a conscious task-based approach to IS&R, however, and one of them is from a different discipline. Yet both studies demonstrate what a task-based approach to IS&R may contribute to IS&R. Harri Laitinen's (1996) study is a M.Sc. thesis in Information Studies investigating the use of an electronic high school learning material collection when answering high school exam questions. D. Isenberg's (1986) study is an investigation of business problem solving and compares experts to novices in the problem solving process and its outcome.

4.3.1. IS&R for high school exams

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

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

The study design contained national exam questions in Finnish high school subjects (religion, physics, biology, psychology, and history), for a total of 12 questions. There were two versions of questions, one set quite verbose hinting at suitable search keys, and the other considerably briefer. The students' task was, using the CD-ROM and search system, to answer one (or more) exam questions within a time allowance of one lesson (45 minutes). The answers were assessed by a teacher using a standard mark scale of 0–6 points (a zero meaning a bad answer). Thus, the assumption was that the use of the electronic collection contributes to the assigned (but very realistic) work task outcome. The study subjects were 29 second year high school students (12 girls, 17 boys), among whom 26 produced notes on their searching and 23 produced answers to one or more exam questions. The system was introduced to the students prior to the test. Multiple data sets were collected: search logs, student interviews, students' search notes, exam question answers, the teacher's assessments and the CD-ROM. Students produced answers to 6 different questions.

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

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

4.3.2. IS&R in business problem solving

Isenberg (1986) studied information seeking and use by experts and novices in business problem solving. His goal was to find out what cognitive processes managers use when they deal with business problems, and about differences in problem solving, and in the quality of the outcome, between experts (experienced managers) and novices (business school students). The experts were 12 experienced general managers from six corporations and the novices were three college undergraduates. All solved the same business case. The case was presented in a controlled way on seven cards, arranged randomly to reflect real business situations where cases do not begin in orderly fashion.

The subjects could utilize the cards as much as they wanted without any extra costs or penalties. Data collection and analysis was through thinking aloud and protocol analysis.

Isenberg found that experienced managers made less use of the information available than novices. Instead, they made more conditional conclusions, which were based on less information supporting them. Experienced managers reasoned from general to specific, based on their experience, and considered several alternatives simultaneously. Powerful deduction and interpretation was typical for them. Isenberg explained this through opportunistic reasoning: if information is valuable but scarce, and its availability unlikely, an expert has learned to distil everything possible out of the information at hand. Therefore the managers made speculative but plausible inferences even on a narrow basis. In the course of problem solving the managers started outlining solutions at an earlier stage than novices and produced qualitatively better results. In other words, experienced managers found a focus (Kuhlthau, 1993) earlier, perhaps directly without any preceding stages, while novices explored each case longer by identifying major concepts and relations before finding a focus. Regarding task complexity, the managers' perceived their tasks as considerably simpler than the novices, with the process, outcome and required information definable at the outset, while the novices had to construct these.

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

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

From an organizational work task perspective IS&R is most often just a supportive function, and its costs in terms of money and effort should be justified by the quality of the deliverables for some necessary and important work tasks. Let us assume therefore that two information access methods (different in some interesting aspect, e.g., collection, IR engine, or interface functionality) are compared from a work task performance perspective in a knowledge intensive task domain. Quality of processes and outcomes, operationalized somehow, are assessed. An interesting difference between methods may be identified either due to its arguable potential contribution to task performance (~technology assessment) or due to the large research efforts it has received in the IS&R community (~a test of ultimate relevance of research efforts). In the following we consider two possible (abstracted) outcomes from the comparison and their consequences.

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

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

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

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

5. DISCUSSION AND CONCLUSION

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

We showed that in studies of IS&R both dependent and independent variables differ considerably. In IR, particularly in experimental studies precision and recall are the dominating dependent variables, whereas in IS the intensity of use of various channels is the most common dependent factor. Users are excluded in Lab IR, but in IS human characteristics are typically independent variables. The independent variables in field studies of IR also differ considerably from those used in IS. These two fields seem to be interested in totally different phenomena. It is very rare to find studies which integrate variables from both camps.

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

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

Scientific theories are needed for the following functions (Bunge, 1967): (i) systematization of knowledge, (ii) guiding research, and (iii) mapping a part of reality. Frameworks and theories of IS&R cannot guide research outside their domains. Since there are relevant phenomena outside their domains, framework/theory enrichment becomes necessary. All enrichment does not nicely fit in the current frameworks but, instead, may require redrawing them.

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

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

There has been much debate about the nature of relevance in IR, and consequently about the dependent variables in IR. In lab experiments recall and precision are exclusively based on topical relevance and not the utility of documents. It is difficult to say whether retrieval by document utility rather than topicality could in any way be supported in document indexing. It is an open question.

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

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

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

Third, task performance, and consequently the information searching it generates, is a process. This means that the search topic, goals and process gradually change when task performance proceeds (Vakkari, 2003). However, there are few studies which have focused on this process.

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

REFERENCES

Alemayehu, N. (2003). Analysing Performance Variation Using Query Expansion. *Journal of the American Society for Information Science and Technology,* 54(5), 379–391.

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

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

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

Belkin, N. J. (1993). Interaction with Texts: Information Retrieval as Information-Seeking Behavior. In: *Information Retrieval '93* (pp. 55–66). Konstanz: Universitetsverlag Konstanz.

Blair, D. (1990). *Language and Representation in Information Retrieval*. Amsterdam: Elsevier.

Borgman, C. L., Hirsh, S. G., and Hiller, J. (1996). Rethinking Online Monitoring Methods for Information Retrieval Systems: From Search Product to Search Process. *Journal of the American Society for Information Science,* 47, 568–583.

Brajnik, G., Mizzarro, S., and Tasso, C. (1996). Evaluating User Interfaces to Information Retrieval Systems. A Case Study. *In Proceedings of the SIGIR'96. ACM, New York,* 128–136.

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

Burgin, R. (1992). Variations in Relevance Judgements and the Evaluation of Retrieval Performance. *Information Processing and Management,* 28(5), 619–627.

Byström, K. (1999). *Task Complexity, Information Types and Information Sources. Examination of Relationships*. Doctoral dissertation. University of Tampere. Acta Universitatis Tamperensis 688.

Byström, K. and Järvelin, K. (1995). Task Complexity Affects Information Seeking and Use. *Information Processing and Management,* 31, 191–213.

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

Chung, Y. and Lee, J. (2004). Optimization of Some Factors Affecting the Performance of Query Expansion. *Information Processing and Management,* 40(6), 891–917.

Cooper, W. S. (1971). A Definition of Relevance for Information Retrieval. *Information Storage and Retrieval,* 7(1), 19–37.

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

Dumais, S., Cuttrell, E., and Chen, H. (2000). Bringing Order to the Web: Automatically Categorizing Search Results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2000* (pp. 145–152). *The Hague.*

Ellis, D. (1989). A Behavioural Approach to Information Retrieval System Design. *Journal of Documentation,* 45, 171–212.

Fowkes, H. and Beaulieu, M. (2000). Interactive Searching Behavior: Okapi Experiment for TREC-8. *In Proceedings of the BCS-IRSG: 22nd Annual Colloquium on Information Retrieval Research* (pp. 47–56). *Cambridge.*

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

Hersh, W., Pentecost, J., and Hickam, D. (1996). A Task-Oriented Approach to Information Retrieval Evaluation. *Journal of the American Society for Information Science*, 47(1), 50–56.

Ingwersen, P. (1996). Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory. *Journal of Documentation,* 52(1), 3–50.

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

Isenberg, D. J. (1986). Thinking and Managing: A Verbal Protocol Analysis of Managerial Problem Solving. *Academy of Management Journal,* 29(4), 775–788.

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

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

Joho, H., Sanderson, M., and Beaulieu, M. (2004). A Study of User Interaction with a Concept-Based Interactive Query Expansion Support Tool. *In Proceedings of the ECIR 2004* (pp. 42–56). *Berlin and Heidelberg: Springer.*

Jones, S., Gatford, M., Robertson, S., Hancock-Beaulieu, M., and Secker, J. (1995). Interactive Thesaurus Navigation: Intelligence Rules OK? *Journal of the American Society for Information Science*, 46(1), 52–59.

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

Koenemann, J. and Belkin, N. (1996). A Case for Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. In*Proceedings of the Human Factors in Computing Systems Conference (CHI'96)* (pp. 205–212). New York, NY: ACM Press Kuhlthau, C. C. (1993). *Seeking Meaning*. Norwood, N.J.: Ablex.

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

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

Laitinen, H. (1996). *Reaalikoevastauksia Rompulta: Tutkimus Elektronisen Oppimateriaalikokoelman Käytöstä Koevastauksien Tuottamisessa [Answering High School Exams with the Help of CD-ROM: An Investigation on Using an Electronic Learning Material Collection When Answering Exam Questions*]. Tampere, Finland: University of Tampere, Dept. of Information Studies, M.Soc.Sc. Thesis, 78 p. [In Finnish]

Morris, P. and Hirst, G. (1991). Lexical Cohesion Computated By Thesaural Relations as an Indicator of the Structure of the Text. *Computational Linguistics, 17*(1), 21–48.

Over, P. (2001). The TREC Interactive Track: An Annotated Bibliography. *Information Processing & Management, 37*(3), 368–381.

Pennanen, M. and Vakkari, P. (2003). Students' Conceptual Structure, Search Process, and Outcome While Preparing a Research Proposal. *Journal of the American Society for Information Science and Technology, 54*(8), 759–770.

Pharo, N. (2004). A New Model of Information Behaviour Based on the Search Situation Transition Schema. *Information Research*, 10(1) chapter 203 [Available at http://InformationR.net/ir/10-1/chapter203.html].

Qiu, Y. and Frei, H. (1993). Concept Based Query Expansion. In: *Proceedings of the ACM SIGIR 16th International Conference*, ACM: New York (pp. 160–169).

Ruthven, I. Lalmas, M. and van Rijsbergen, K. (2003). Incorporating User Search Behavior into Relevance Feedback. *Journal of the American Society for Information Science and Technology,* 54(6), 529–549.

Saracevic, T. (1976). Relevance in Information Science. In: M. Voigt and M. Harris (Eds.) *Advances in Librarianship*, vol. 6. Academic Press: New York (pp. 79–138).

Saracevic, T. and Kantor, P. (1988). A Study of Information Seeking and Retrieving. Part II: Users, Questions, and Effectiveness. *Journal of the American Society for Information Science,* 39, 177–176.

Schamber, L. (1994). Relevance and Information Behavior. In: M. Williams (Ed.), *Annual Review of Information Science and Technology,* 29 (pp. 3–48). Medford, NJ: Information Today.

Sihvonen, A. and Vakkari, P. (2004). Subject Knowledge Improves Interactive Query Expansion Assisted by a Thesaurus. *Journal of Documentation,* 60(6), 673–690.

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

Stairmand, M. (1997). Textual Context Analysis for Information Retrieval. In: *Proceedings of the ACM SIGIR 20th International conference*, ACM: New York, 140–147

Stokes, N. (2004). *Applications of Lexical Cohesion Analysis in the Topic Detection and Tracking Domain*. Doctoral Dissertation. National University of Ireland, Dublin.

Stubbs, M. (2001). *Words and Phrases. Corpus Studies in Lexical Semantics*. Blackwells: Oxford.

Suomela, S. and Kekäläinen, J. (2005). Ontology as Search Tool: A Study of Real Users' Query Formulation with and Without Conceptual Support. In: D. E. Losada (Ed.), *Proceedings of the 27th European Conference on Information Retrieval, 21–23 March, 2005, Santiago de Compostella*.

Vakkari, P. (1997). Information Seeking in Context: A Challenging Metatheory. In: P. Vakkari, R. Savolainen and B. Dervin (Eds.) Information Seeking in Context. In *Proceedings of an International Conference on Research on Information Needs, Seeking and Use in Different Contexts* (pp. 451–464), 14–16 August 1996, Tampere, Finland. London: Taylor Graham.

Vakkari, P. (1998). Growth of Theories on Information Seeking: An Analysis of Growth of a Theoretical Research Program on the Relation Between Task Complexity and Information Seeking. *Information Processing and Management,* 34, 361–382.

Vakkari, P. (2001). A Theory of the Task-Based Information Retrieval. *Journal of Documentation,* 57(1), 44–60.

Vakkari, P. (2003). Task Based Information Searching. In: B. Cronin (Ed.) *Annual Review of Information Science and Technology:* Vol. 37 (pp. 413–464). Medford, NJ: Information Today.

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

Vakkari, P. and Sormunen, E. (2004) The Influence of Relevance Levels on the Effectiveness of Interactive Information Retrieval. *Journal of the American Society for Information Science*, 55, 963–969.

Vorhees, E. (2000), Variations in Relevance Judgements and the Measurement of Retrieval Effectiveness. *Information Processing and Management,* 36, 697–716.

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

Wagner, D. and Berger, J. (1985). Do Sociological Theories Grow? *American Journal of Sociology,* 90, 697–728.

Wang, P. and Soergel, D. (1998). A Cognitive Model of Document Use During a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science,* 49, 115–133.

Wersig, G. (1973). *Informationssoziologie: Hinweise Zu Einem Informationswissen-Schaftlichen Teilbereich*. Frankfurt, BRD: Athenäum Fischer.

CHAPTER 8

TOWARDS AN ALTERNATIVE INFORMATION RETRIEVAL SYSTEM FOR CHILDREN

JAMSHID BEHESHTI, LEANNE BOWLER, ANDREW LARGE, AND VALERIE NESSET
McGill University

1. INTRODUCTION

A recent survey of more than 1700 households indicates that the main reason many parents purchase computers and connect their children to the Internet at home is for education (*Safe and Smart*). In addition the survey shows that children also use the Internet for educational activities that go beyond required school work. In fact, the fastest growing group of Internet users are children between the ages of eight and twelve (Vise, 2003), who are increasingly using the Web to access educational as well as entertainment materials. Children, however, rely on conventional information retrieval (IR) systems and search engines intended for general adult use, such as MSN or Google, and to a much lesser extent, Web portals such as Yahooligans! and LycosZone specifically intended for young users (Large et al., 2004; Large, Beheshti, and Rahman, 2002a).

But research has shown that children's information needs (Walter, 1994), research approaches (Kuhlthau, 1991), and cognitive abilities and higher order thinking skills (Neuman, 1995; Siegler, 1998; Vandergrift, 1989) differ from those of adults. The results of earlier studies on children's use of online catalogues designed for adults indicate that young users are often faced with difficulties locating specific information related to their information needs (Hirsh, 1997). A growing body of research points to the problems children typically encounter when seeking information on the Web. Kafai and Bates (1997) conducted one of the first studies with young children on their use of Web sites, and concluded that they were able to navigate through the links and scroll. Only the older children, however, could use search engines effectively.

Hirsh (1999) investigated the searching behavior of ten fifth graders and concluded that they encountered difficulties in formulating effective search queries and did not use advanced features. Schacter, Chung, and Dorr (1998) conducted a study on Internet searching by fifth and sixth graders and concluded that they did not plan their searches, used ill-defined queries, and preferred browsing. Large, Beheshti, and Moukdad (1999), investigating the information seeking behavior of 53 sixth graders, similarly found that children preferred browsing to searching. Bowler, Large, and Rejskind (2001), focusing on a few case studies of grade six students concluded that search engines designed for adults are unsuitable for children. Wallace et al. (2000), studying sixth graders, discovered that experience in using search engines does not improve children's search strategies and in general information seeking is an unfamiliar activity for children.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 139–165.

This chapter focuses on the theoretical and applied research conducted in several disciplines as the foundation for a new information retrieval system consisting of conventional approaches to retrieving information along with a virtual reality environment, through which children can browse. The chapter is divided into five sections. Section 1 is the Introduction. In Section 2, we begin at the broadest level, describing the four dominant frameworks for learning research. Each of these frameworks is based on a metaphor of how the human mind works when learning. The current dominant metaphor is social constructivism (collaborative IR), which must be reconciled with the individual nature of much of IR searching (the cognitive constructivism metaphor). In Section 3, we describe the affective factor when modeling children's search processes, basing it on Kuhlthau's (1993) Information Search Process (ISP) model. We build into the model affective, empathetic and intermediary or tutoring dimensions, which must be considered for new IR systems for children. In Section 4, we highlight browsing features that correspond to children's searching behavior during the intial stages of their ISP. In Section 5, we propose an alternative IR system for children based on the previous issues and discussion.

1.1. Children

The term "children" may be applied to a wide range of ages, from toddlers to high school seniors. Piaget, from a structuralist perspective, suggested that cognitive developments occur in stages based on age bands of birth to two years, two to seven, seven to 11, and 11 to 16 years (Piaget and Inhelder, 1969). He argued that children need to experience these stages of cognitive development and that younger children do not have the capacity or capability to think in the same way as older ones. Bandura, on the other hand, used a behavioral approach suggesting that children learn through imitating role models and observing their environment (Bandura and Huston, 1961). Whereas Piaget believed that children's major cognitive development takes place before the age of 12, Bandura suggested that development occurs throughout life and is not confined to one or more stages.

In information-seeking behavior age-based categorization may be somewhat artificial, where it might be difficult to distinguish one group from another as some similarities have been observed between the groups. Fidel et al. (1999) investigated the information-seeking behavior of eight grade 11 and 12 students and found that they, too, needed formal training to be effective Web searchers. Agosto (2002) interviewed 32 grades nine and ten students and concluded that they may also encounter problems interacting with IR systems and that a better understanding of their decision-making processes is needed to assist them in searching.

Based on an extensive review of the literature, Large (2005) comments that only one study by Bilal and Kirby (2002) has been published that explicitly investigates the differences between children and adults in searching for information on the Web. Bilal, and Kirby (2002) studied the cognitive, affective, and physical behavior of 14 seventh graders and nine graduate students in information science. They utilized a number of metrics to measure the two groups' effectiveness and efficiency in using Yahooligans!, when given specific tasks to search. They found that whilst the graduate students had a success rate of 89 percent, the seventh graders only managed a 50 percent success rate. The latter group also frequently revisited sites and took more time in finding information.

Large's (2005) literature survey leads him to conclude that the verdict is still out regarding the differences and similarities in IR behavior of various age groups. What has been reported in the literature, Large (2005) states, demonstrates that children encounter problems in selecting appropriate search terms, move quickly through the Web pages spending little time reading the materials, and have difficulty judging the relevance of the retrieved pages.

While only a few studies have been undertaken on the differences between children and adults, we can safely assume that in the context of information seeking and retrieval, whether in a traditional library or on the Web, both groups' cognitive, affective, and physical behavior may contribute to their success or failure (Bilal and Kirby, 2002; Kuhlthau, 1993). For the purpose of this chapter, we have defined "children" in a broad sense covering grades one through senior high school, and we have focused our attention on the cognitive and affective factors, followed by a discussion on an alternative IR system for children.

2. METAPHORS OF THE MIND

The Web is a "vast collection of completely uncontrolled heterogeneous documents" (Brin and Page, 1997, Section 3.2) and within this collection, processes occur that bring together information and information seekers.

2.1. The Information Processing Metaphor

How information science views these processes are shaped by a guiding metaphor of the mind and a vision of how humans understand their world. A dominant metaphor in information science, borrowed from the domain of cognitive psychology, has been the information processing approach to human cognition. As a driving theme in information science it has influenced the way we study information behavior and design IR products for users. Information processing theory uses the computer as a metaphor for how humans interface with information and views the mind as an information-processing system. In this view, human information behavior is rule-based; information is input, processed and retrieved according to a model or plan of how the world fits together. The more complete the model, the better information can be retrieved. An assumption of this metaphor of the mind is that there is an external truth, a specific model of knowledge that can be replicated by a computer system. This assumption has led to studies and research into information systems built upon the mental models of experts, that is to say, systems that model knowledge based on how experts in a field view the problem space. Taxonomies, indexes and thesauri also try to model the "perfect" knowledge of experts through semantic structures.

2.2. The Constructivist Metaphor

Information-processing theories have been called the "second metaphor" for learning, the first being Behaviorism (Mayer, 1997). More recently a third metaphor has arisen in cognitive psychology. The central tenant of this paradigm for "knowing" is that humans construct their own knowledge. In this view learning is a constructive process rooted in authentic tasks. As cognitive theory has shifted toward constructivism, so has information science.

Constructivism in information science is reflected in the growing awareness of the importance of a user-centered approach to the design of information systems and services (Nahl, 1995; Saracevic, 1999). A related body of research views information seeking as a sense-making, knowledge-constructing process unique to each individual and situation (Belkin, 1980; Dervin, 1999; Kuhlthau, 1991, 2004). Information scientists have begun speaking of processes that evolve over time. One example of this is Kuhlthau's model of the information search process which showed that people's conceptions of information change as they progress through the information problem (Kuhlthau, 1991, 2004). The concept of information literacy, an educational outcome applauded by a growing number of teaching institutions, has evolved from the constructivist metaphor of learning. An information literate person is able to find, evaluate and use information in ways that are appropriate to the need. The task is cognitive, but is essentially constructive in that it is an evaluative process and is linked to information needs situated in authentic tasks.

2.3. The Social Constructivist Metaphor

And now a fourth metaphor is arising "under the banner of social constructivism" which views learning as "social negotiation and learners as social negotiators" (Mayer, 1997, p. 154). The new metaphor incorporates all that traditional constructivism represents and more. Traditional constructivism approached learning and knowing from a strictly cognitive perspective. The fourth metaphor extends the metaphor into areas that include affective, social and cultural aspects of knowledge construction. With this view we see a shift from the individual and his/her mind to a focus on a community of learners within which the individual operates. A more holistic approach, this view acknowledges that thinking and learning do not happen in a box—they are related to feelings, motivation, local culture and even biological and physiological imperatives.

Within the social constructivist framework, the conception of childhood changes as well. In the Social Constructivist paradigm, children are not empty vessels waiting to be filled with knowledge by adults. Rather, children are complex, social beings who, as agents of their own learning, need IR tools that will assist them in ways that make sense to them.

2.4. Toward the Fourth Metaphor: Re-aligning Information Science

The information processing metaphor has held court in information science for many years and in some ways it has been a comfortable fit—information systems, in this view, could be neatly bundled to reflect the expert knowledge of indexers and computer programmers who wrote algorithms for matching search terms to information resources. Information seekers, once their own mental models matched those of the experts, could, it was assumed, negotiate through the universe of information with ease: designers of information systems need not concern themselves with external forces, such as the user's motivation, anxiety or influences from the larger society.

As Social Constructivism becomes a predominant metaphor of the mind, we may see the incorporation of social aspects into IR. This may create some confusion for information science as it re-aligns its guiding metaphor. There will be difficulties in this task. As one example, applying the concept of a Zone of Proximal Development—a social constructivist concept—to information seeking, an activity that is essentially conducted alone, and most often with the use of an electronic data base, suggests a conflict in philosophy and values.

2.5. The Zone of Proximal Development: IR from a Social Constructivist View

At the heart of the concept of a Zone of Proximal Development is a sociocultural approach to knowledge development. First articulated by the Russian psychologist Lev Vygotsky in the 1930's, the concept has become a driving force in the field of education, particularly early childhood and elementary education (Vygotsky, 1978). What is the Zone of Proximal Development? Simply put, it is the bandwidth of competence within which people learn. At the lower limit of the zone are the tasks that a person can accomplish independently, while at the upper limit is the space where more complex tasks can be realized by a person through interactions with more knowledgeable others.

2.6. Communication: The Key to the Zone of Proximal Development

Communication and social interaction are key features of the Zone of Proximal Development. Knowledge-building is co-operatively achieved through scaffolding, the supportive framework that is adjusted and modified by the expert participant (the teacher or the coach) according to the novice's needs. Kuhlthau's notion of the *Zone of Intervention* in reference and library instruction is modeled upon Vygotsky's Zone of Proximal Development and is defined as "that area in which an information user can do with advice and assistance what he or she cannot do alone or can do only with great difficulty" (Kuhlthau, 2005; see particularly section entitled: *Design Application for Implementation into Practice*). Cooper (2002), in her study into the information-seeking behavior of seven-year old children, makes mention of both Vygotsky's Zone of Proximal Development and Kuhlthau's Zone of Intervention, suggesting that when searching in supportive contexts, children are indeed capable of moving from non-specific browsing strategies to more focused searching strategies.

2.7. Supportive Contexts

The notion of a supportive context is central to the Zone of Proximal Development and a functional Zone of Proximal Development implies dialogue with other people. Learning is situated in a social context. In the classroom setting, the supportive context is the teacher and the other students, working together in a collective fashion to create a community of practice. The teacher's role in the supportive context is crucial, as he or she provides the link between the learner and the knowledge to be learned. Teachers facilitate the internalization of knowledge by confining classroom activity to the Zone of Proximal Development.

This supportive context can apply to the library environment as well. In the Cooper (2002) study of seven-year old children using the library, it took only a small show of encouragement or a well-timed question-prompt from an adult for the children to use advanced search techniques—nevertheless, dialogue was crucial to the intervention. For information science, the concept of dialogue can only make sense if IR is viewed as a necessarily collaborative effort. But how do you collaborate with an IR system?

Researchers in the domain of educational psychology speak of developing Zones of Proximal Development electronically, through the creation of communities of discourse, which may exist through the medium of communication technologies such as e-mail, online discussions and chat-rooms (Brown, Ellery, and Campione, 1998). How can the concept of a community of discourse be transferred into information technology, where individuals seemingly work in solitude while they search for information? This suggests a radical overturn of the way we think about IR, particularly for children.

2.8. A Paradox for Information Science: IR as a Community Activity?

In the context of Web-based IR, the Zone of Proximal Development is problematic. Given that the search process involves learning from a variety of inconsistent and often incompatible sources, and that information seeking is often, if not usually, performed independently, without the guidance of teachers or librarians, we might ask what happens when children attempt to solve information problems on their own? Without the support of human mediation, how do students progress through the information problem solving process? Should we be designing information systems that encourage discourse? Should we be building tools that facilitate the development of communities of learners? Should information systems be designed that shift methods to suit the social/cultural/affective context of the user? Kuhlthau (1999) has fine-tuned these questions by posing her own challenge: Can IR systems be designed to provide the kinds of support needed during each zone or developmental stage of information problem solving?

2.9. Search Pals for Children

Web-based IR is usually a solitary activity, even for children, and finding a supportive context is problematic if not paradoxical. Perhaps there are lessons to be learned from other disciplines.

2.10. Pedagogical Agents

From the field of intelligent tutoring systems comes the idea of pedagogical agents—autonomous agents that facilitate learning through interactions with students (Clarebout et al., 2002; Moundridou and Virvou, 2002; Moreno et al., 2001; Shaw, Johnson, and Ganeshan, 1999). A recent development in intelligent tutoring systems, pedagogical agents are animated characters designed to support learning. The work of these agents is not to transmit knowledge—they do not simply pass on information or answer fact-based questions. Rather their role is to act as facilitators in the knowledge building process. Working in the Zone of Proximal Development, they facilitate knowledge building by scaffolding, modeling, demonstrating, coaching, testing and performing tasks too difficult for the learner to perform, all within a reciprocal communication between agent and learner (Clarebout et al., 2002).

Interestingly, pedagogical agents can exhibit personality and emotions which seem to enhance student motivation and positive perceptions about the learning experience. (Moundridou and Virvou, 2002). Some examples: *Herman the Bug* is an animated character who inhabits a learning environment called "Design-A-Plant". His role is to assist students in a problem-solving task related to the domain of botanical anatomy and physiology through "just in time" support communicated verbally and non-verbally (Moreno et al., 2001). *Adele* is a female animated "student advisor". Using a combination of explanation and question prompts, she supports distance learning in the domains of medicine and dentistry (Shaw, Johnson, and Ganeshan, 1999).

2.11. Search Agents for Children

Could a similar type of agent be applied to an IR system for children? We might call it a "search agent" or even (for children) a "search pal". Clarebout et al. (2002) speak of the need for pedagogical agents for open-learning environments due to the lack of support tools in such environments. Passive tools such as conventional help screens, however, are seldom used, a finding confirmed by Large, Beheshti, and Breuleux (1998) in their study of children searching a CD-ROM encyclopedia. The children in this study rarely sought help, and did so halfheartedly.

We can only speculate as to how a search agent might help children retrieve information from the Web. Pedagogical agents have been designed to provide support within a specific and narrow domain of knowledge; indeed none of the pedagogical agents described in the literature could function on their own outside of their content area. How would such an agent function in the complex, open-ended, heterogeneous environment of the Web?

Clarebout et al. (2002) suggest a different approach: provide metacognitive support instead of domain-specific support. This idea is compatible with Kuhlthau's (1997) suggestion that learning environments should be created in digital libraries based on constructivist approaches. The Web should, according to this view, be seen as more than an information-gathering tool. Rather it should be seen as a constructivist tool for helping users find personal meaning in information. A focus on metacognitive processes might help young users find meaning by providing the kind of support that children need to help them monitor their own progress through the information-seeking process. Knowing when to ask for help and then receiving it in a timely fashion, all within a positive, social context, might be what children need to sustain them in the complex, sometimes troublesome world of the Web.

3. THE AFFECTIVE FACTOR

Retrieving information from the Web is as much an emotional as an intellectual experience. This may be all the more true for young people, who, as novices in all aspects of life, might conceivably experience the Web in different emotional terms than adults. As information seekers, children face information problems with undeveloped reading and analysis skills and little context in which to locate and understand new information, a situation with great potential for disorientation and feelings of confusion and frustration.

Research in information-seeking behavior has explored information behavior primarily within a cognitive framework. Increasingly, however, information seeking is seen as a multi-dimensional process; an interplay between cognition, affect and actions. Calls for a broader perspective have shifted the focus to a holistic perspective of human information behavior that includes an understanding of emotion, mood, motivation, and preferences (Nahl, 1995; Wilson, 1981, 1997). Even so, affective issues continue to receive minimal attention from authors in information-systems related research in the area of LIS (Julien, McKechnie, and Hart, 2004) and even less so in the area of children, the Web and IR. The number of researchers who have investigated the effect of children's attitudes, emotions and motivations on IR in the Web-based environment is indeed small, suggesting a potential area for further research (Bilal, 2000; Bilal, 2002b; Bilal and Kirby, 2002; Large et al., 2004; Large, Beheshti, and Rahman, 2002a; Tsai, Lin, and Tsai, 2001; Watson, 1998).

3.1. Information Search Process Model: The Affective Dimension

Youth and the role of affect in information-seeking were first studied by Kuhlthau in her groundbreaking work on the Information Search Process (Kuhlthau, 1991, 2004). Although this series of studies did not investigate information seeking within the framework of an open-ended information system such as the Web, the resulting model provides a useful lens for further analysis into children, affect and the Web.

3.2. The ISP Model: A Three-Dimensional View of Search Behavior

Kuhlthau charted the course of high school students' feelings along the projectory of an information task—completing a research project for school—and found a correlation between the degree of understanding that students had of the topic space, their level of anxiety, and their ability to complete the information task successfully. From these studies emerged the Information Search Process (ISP) model, a three dimensional model incorporating three modes of behavior—cognition, affect and physical actions. In the early stages of the information search process, before a focus (or understanding) has been identified by the user, high levels of uncertainty are typically reported. Finding the focus is the turning point in the process, where feelings of uncertainty are substituted with feelings of optimism.

3.3. Uncertainty

Although it may seem counter-intuitive, uncertainty felt during the early stages of the information search process does not necessarily represent a road-block or a dead end in the search process. Rather, uncertainty indicates a potential area of learning, a space where users can begin to accommodate a new understanding of the topic area they are researching. The feelings of anxiety associated with uncertainty can, however, lead information seekers to abandon their information search if the feelings persist. It seems, then, that alleviating the cognitive state of uncertainty might create affective conditions of confidence that will help information seekers move forward in the process.

3.4. Children, Affect and the ISP

Little research has focused on children, affect and the ISP. Cooper's (2002) study of seven-year old children seeking information provides a glimpse into this relationship, finding connections between the children's lack of self-confidence and their ability to search effectively. Although this study did not look specifically at children using the Web, inferences can be made concerning the importance of affective considerations in the design of information services for young children.

Cooper found that, although certainly capable of using the meta-data found in electronic encyclopedias, the children were hesitant to do so when they worked independently of a more knowledgeable guide. Lacking the self-confidence to take "risks" the children instead chose to continue searching the safe way, using browsing strategies rather than analytical searching. Only when guided and encouraged by someone more knowledgeable were they able to focus their search strategies. Cooper emphasizes the importance of feelings during the search process for novice information-seekers.

3.5. Children and the Web: Potential Effects on the ISP

Despite the impact of the ISP model in the field of LIS, very little research directly related to the model and children in the context of the Web has been reported in the literature. There has been speculation as to how the ISP model would play out in the context of an IR system (Kuhlthau, 1997) and children's search engines on the Web (Broch, 2000). The Web presumably magnifies the affective problems created by uncertainty by virtue of the greater difficulty achieving a sense of focus in such an over-abundant, open-ended information space. A pitfall for children when using the Web may be feelings of being overwhelmed by the sheer volume of information. Stuck in a perpetual loop, children could remain at the high anxiety end of the ISP, repeating the initial stages over and over again until they simply give up in frustration.

3.6. Children's Attitudes Toward the Web

IR studies that have investigated the role of affect on children searching the Web have focused on broader attitudinal aspects related to the novelty or interest, rather than, as in the ISP model, anxiety specifically related to a particular stage in the information seeking process or the information seeker's understanding of the topic.

In one of the earliest studies in the area of children's use of the Web, Watson (1998) used qualitative methods to reveal eighth-grade student's experiences and reflections on using technology as a tool. The students had access to a variety of information technology, including CD-ROM drives, laser-disc players and Internet connectivity. Watson found that, despite its difficulties, children exhibited positive attitudes and self-confidence in relation to their use of information technology, and in particular, the Internet, although positive feelings did not necessarily indicate the successful completion of a search task.

Bilal's three-part series of studies of seventh-grade children searching the Web using the Yahooligans! search engine/directory confirmed Watson's findings (Bilal, 2000, 2001, 2002b). Bilal (2002b) recorded the specific Web moves used to locate information rather than the broader stages in the process. An interview followed the search sessions in order to gauge attitudes and perceptions. The children were asked to complete three different types of search tasks; an assigned fact-finding task on a science topic, and two complex tasks that required the interpretation of meaning—one an assigned research-oriented task and the other a self-generated task.

In Bilal's study, a high percentage (87 percent) of the children had positive feelings toward searching the Web, saying that they enjoyed using it over other types of information sources because of; "(a) ease of use over other types of sources, especially print; (b) ability to employ keyword searching; (c) visiting different Web sites to find the information; (d) availability of graphics, and (e) fun (Bilal, 2000). The children reported few negative feelings, but those who did related it to the "difficulty in finding the answer" and a "lack of matches" (p. 659).

Most children in Bilal's study expressed strong motivation; Yahooligans! apparently was able to provide interesting content, intellectual challenge and convenient access from home. Despite the warnings expressed by Broch (2000) regarding feelings of being overwhelmed by information when searching the Web, the children in Bilal's study were persistent and patient information seekers when answering a fact-based question. The reasons for their resilience were that the Web allowed for "efficiency, exploration, challenge, and convenience." (2000, p. 659). Interestingly, most children (47%) reportedly preferred the self-generated task, with a full 73% succeeding in this task, a finding that supports educational theory about the positive role of individual interests in processing information (Blumenfeld et al., 1991; Hidi, 1990). Bilal warns, however, that the children often chose topics that were very broad in nature and that mediation was needed to help the children narrow the focus of their information need.

In the belief that an affective attribute can effect motivation and performance, Tsai, Lin, and Tsai (2001) created an Internet Attitude Scale for Taiwan high school students to measure their levels of anxiety and confidence as well as their perception of the Internet's usefulness. The Internet Attitude Scale ranked perceived usefulness, affection, perceived control, and behavior.

The study also explored gender differences and prior experiences using the Internet. The study reports that, irrespective of gender or experience, students in general appreciated the usefulness of the Internet and showed positive feelings toward it. Gender differences did emerge when factors such as anxiety and confidence were analyzed; the males expressing lower anxiety and higher confidence levels than the female students (2001). Attitudes were not related to the students' ability to successfully complete a search task.

Large, Beheshti, and Rahman's (2002b) study of grade-six boys and girls searching the Web to complete a school assignment also showed gender differences in affective responses to the Web. As with Bilal, interaction with the Web was analyzed within the framework of specific moves on the screen, such as, for example, keyword/browsing search strategies or time spent viewing Web pages. While both genders assessed information on the Web using affective criteria (novelty and interesting information) rather than meaningfulness, the boys seemed to be less engaged with the content, spending less time viewing and more time clicking hyperlinks.

3.7. Attitudes and Outcomes

Although most of the research on children, affect and the Web seems thus far to indicate strong positive attitudes, it should be noted that first, there has been little linkage between general impressions/perceptions and final outcomes (ability to solve the information problem) and secondly, much of this research was conducted nearly five years ago, when the Web was still a novelty to many children. The paucity of research investigating children's affective behavior in the realm of Web-based IR suggests an area for further investigation.

All of the studies described above focused on middle- and high-school students. Yet it is known that youngsters in the early years of elementary school do use the Web to find information (Large, 2005). Further study into how children of this age react to the Web as an IR tool should be undertaken. Another area to explore is the nature of the information task vis à vis the Web. Bilal's (2000, 2001, 2002b, 2002c) studies suggested attitudinal differences between tasks imposed and tasks that are self-generated. What sorts of affective behaviors do children have when they retrieve information on the Web in the context of a personal information need that is not related to a school assignment?

3.8. Empathetic IR Systems for Children

Thus far discussion has focused on *children's affective reactions to the Web*, but what about the *Web's affective reaction to children*? Can we speak of empathetic IR systems; information systems that mediate IR via affect? Given that retrieving information from the Web most likely occurs without intermediaries, can we design IR systems that respond to the emotional and motivational needs of young information seekers in appropriate and timely ways?

3.9. Intelligent Agents

Information science has tried to play the facilitator role in IR through the design of Intelligent Agents, software agents that act as "electronic butlers", helping users perform routine tasks (Baylor, 1999; Negroponte, 1997; Perez, 2002). Research in the field of Educational Technology is extending the "butler" metaphor toward one of "coach" or, intelligent cognitive tool, augmenting a person's zone of proximal development. Typically, these "coaches" diagnose problems from a cognitive point of view, by providing ways to solve the problem (effective search strategies) or advanced mental models of a domain of knowledge (ways to understand the problem space) (Baylor, 1999). But children, especially young children, may need more that pure cognitive support when they search the Web. Encouragement, enthusiasm, and an attitude of good humor and fun may be equally valuable to young information seekers lacking the confidence and experience to navigate the complex world of the Web. A socially-intelligent search agent that emotes as it "coaches" information seekers may be preferable in the world of the child.

A look at the literature from the field of Educational Psychology might reveal what such a socially-intelligent agent might look like in the context of an IR system for children. In their article, *Socializing the intelligent tutor: Bringing empathy to computer tutors,* Lepper and Chabay (1988) ask the deceptively simple question, can empathy be built into a computer tutor (or "coach" as the case may be)? (Mandl and Lesgold, 1988).

To answer this question, Lepper and Chabay first look at the motivational goals of human tutors and how these goals can be transferred to the intelligent tutoring system. First among these goals is the desire to keep students from "becoming so discouraged, frustrated, or alienated that they give up on the task at hand." To do so, tutors will "reassure, commiserate, exhort, and encourage." Tutors also seek to encourage high levels of attention and effort and will "goad, provoke, or reprove their pupils" (p. 247). Such feedback, based on the tutor's assessment of the student's state of mind, helps to fuel learning that is extrinsically motivated.

Lepper and Chabay also identify ways that tutors can promote intrinsic motivation by enhancing feelings of challenge, curiosity, or control. Challenge can be addressed by setting appropriate performance goals. Curiosity can be maintained through the placement of well-timed hints or clues to help solve a problem. A sense of learner control can be developed by a deliberate choice on the part of the tutor to withhold intervention, permitting learners to make mistakes (p. 247).

Can the motivational goals of tutors be transferred to the IR environment? Rephrased, this question becomes, can an IR system provide users with the kinds of motivational incentives used by human tutors? To answer this question, we must find a space where the task of the tutor and the task of the IR system overlap. This is possible if we think of IR from the constructivist point of view. In this conception, the information professional is more than an information organizer and gatherer—she is a facilitator and counselor, a partner in the construction of knowledge. The goal of the information professional is to provide support in the sense-making process, in much the same way that a human tutor might provide support for a learning activity. As agents of the information professional, IR systems should in theory fill this same facilitative role.

3.10. Applying the Motivational Goals of Tutors to the IR system

We generally speak of two forms of motivation—extrinsic and intrinsic. Extrinsic motivation is driven by reinforcement external to the individual. Prizes, grades, verbal encouragement from tutors and coaches are all incentives that reward children who are driven by extrinsic motivation. Intrinsic motivation does not require such rewards and is driven instead by a child's internal goals—tasks are completed because the task is itself pleasurable and leads to feelings of success. IR systems should try to promote both.

3.11. Extrinsic Motivation

Although motivation that is driven by intrinsic orientations is more powerful and long lasting, there are times when promoting extrinsic motivation is "necessary in order to produce learning when the activity is one that students do not find of inherent interest or value; extrinsic motivations thus may actually produce positive effects on learning when tasks are of low initial value" (Lepper, 1988, p. 299). This is certainly true for information seeking, which is often a task imposed on the user by an external source.

In computer-based environments, extrinsic motivation takes the form of system feedback resulting from the actions taken by the user. The classic example of feedback that does *not* motivate is Microsoft's warning to users that they have performed "an illegal action", causing novice computer users to think they've broken the law. Socially intelligent search agents can overcome this negativity by paying attention to general guidelines concerning the appropriateness of various social and motivational remarks in different situations. For example, a socially intelligent tutor should offer reassurances, rather than negative responses. The same principle could apply to an IR system. Information seekers in the exploration stages should receive messages that encourage discovery. Remarks that imply incompetence on the part of the searcher are not helpful motivators (like *KidsClick*'s enormous message in bold that reads, "SEARCH ERROR").

3.12. Intrinsic Motivation

System feedback from socially intelligent search agents can also promote intrinsic motivation by modeling the ways that human tutors provoke curiosity, provide challenges and develop a sense of control in the learner. Finding the right balance between control and curiosity on the one hand, and challenge on the other, can be tricky. Too much intervention will reduce a child's sense of control and innate curiosity while too little may make the task too complex. To find the right recipe the socially intelligent search agent must create a diagnostic model of the information seeker. This would be done by observing the errors made during the search task or by direct questioning of the user by the agent. This suggests a much more interactive approach to IR system design.

Imagine a system that diagnoses where a user is in the search process. Such a system has been proposed for undergraduate students, where a navigation classification scheme uses the metaphor of a ship and lighthouses to guide the user through 'charted' waters of information (Leide et al., 2003). The lighthouses contain mediation interfaces with agents to link the user to information. Each agent is customized after the user's cognitive model of the information space to interact with the system's agent.

Cole has also called for an "enabling" IR system, where the system actively and interactively helps the user to understand better the information need, task, and the problem, through a decoding and encoding operation (Cole, 1999, 2000). Conceivably, an IR system that "knows" a user is in the preliminary stages of the search will also know that it is crucial for them to explore, discover, take some risks, make mistakes—responses rarely, if ever, seen in current IR systems. It will also know when the task becomes too frustrating and will offer context-appropriate responses that will guide the user, without diminishing the feelings of curiosity, challenge and control.

3.13. Modeling Affect and Cognition

The socially intelligent search agent, like human tutors, might have special background knowledge of the information seekers, such as their prior knowledge, their learning style, aptitudes and motivations. This suggests that IR systems should be customized to match the needs of each user. Conceivably one day every information seeker might have residing on a personal computer, a personalized system that "understands" the ways, both cognitively and emotionally, that they look for information. To do so will require the preliminary modeling of the affective and cognitive responses of human information seekers.

Purists in the field of Artificial Intelligence and Human Computer Interaction would base computer responses to human affect on the user's physical expressions of emotion, such as changes in biochemical and electrical signals or visual cues from facial expressions, because "the sine qua non of emotion expression is the physical body" (Picard, 2003, p. 6). Issues of privacy aside, this method, though perhaps useful for mapping out feelings in the broadest sense, does not provide a finely tuned image of what people are thinking and feeling. More useful is the theory emerging from the field of Human Information Behavior as, for example, Kuhlthau's series of studies on the ISP, which showed a correlation between feelings, the degree of focus and the ability to complete the information task successfully (Kuhlthau, 1991, 2004). More in-depth studies involving Kuhlthau's and other information behavior models may help designers of Web-based IR systems create the information structures needed to address the user's point of view in the search process.

Mapping out the feelings experienced during IR can provide us with clues as to the affective barriers faced by children seeking information and is a useful way to predict *when* it might be useful for facilitators (of either the human or electronic variety) to provide motivational incentives.

Motivation is not the same attribute as feelings: however, motivation can be affected by feelings. In the case of Kuhlthau's model, the dip in feelings of confidence and interest, accompanied by a rise in feelings of frustration and confusion that were experienced during the exploratory/discovery stage of the ISP resulted in a situation of low motivation. This then is the space where motivational incentives should be targeted.

4. BROWSING DURING THE INITIAL STAGES OF ISP

> I find that a great part of the information I have was acquired by looking up something and finding something else on the way.*

More than three decades ago, Apted (1971), Herner (1970), Hyman (1971) and Lancaster (1968) provided several definitions of and categories for browsing in relation to traditional libraries, which were summarized and labeled by Beheshti (1992) as random or systematic. While random represents the non-purposive, general type of browsing mainly for interest and leisurely reading that may occur in public libraries, systematic browsing may have a specific purpose, namely to alleviate the user's information uncertainty. At this stage, the user "may have a need which is so ambiguous that no search specification is possible. Instead, the bookshelves or database are scanned in

---

* Franklin P. Adams (1881–1960) was an American journalist and humorist who wrote numerous articles and books. Roberts, N. L. (1984). Franklin P. Adams. In: *American Newspaper Journalists Dictionary of Literary Biography*, 29, p. 18.

order to explore possible matches between the intuitive current need and the available items." (Pejtersen, 1989, p. 41).

Other browsing typologies have been reported in the literature (Chang and Rice, 1993; Hildreth, 1987), which may lead to the conclusion that browsing occurs on a spectrum of information seeking behaviors. At one end of the spectrum, a tacit information need does not exist—people simply browse for pleasure, whilst at the other end of the spectrum the user is browsing a highly organized list to retrieve a known item. The wide range of activities may be summarized in this recent definition of browsing (Rice, McCreadie, and Chang, 2001, p. 302),

> Browsing, in essence, is an examination of unknown items of potential interest by scanning or moving through an information space in order to judge the utility of the items, to learn about something of interest in the item, or to satisfy curiosity about something.

The underlying assumption is that browsing is a "natural and effective approach to many types of information-seeking problems," which "requires a smaller cognitive load than analytical search strategies: do" (Marchionini, 1995, p.103). Marchionini distinguishes between different browsing strategies: scanning, observing, navigating, and monitoring. Scanning is a perceptual recognition or matching activity, where a well-defined object in the user's mind is compared to a list of objects. Scanning requires low cognitive effort, but a highly organized environment. Observational strategy requires interpretation to make sense of what is observed and as such it needs a higher cognitive load. Navigation occurs in environments where the user is presented with different routes and options to choose from, requiring a relatively high cognitive load. The monitor strategy is similar to scanning, entailing little cognitive load, but is not dependent on environmental stimuli. Marchionini concludes that information systems should accommodate all these types of browsing strategies.

Although browsing has been a prevalent information-seeking behavior, very little is known about what motivates people to choose this strategy (Rice, McCreadie, and Chang, 2001). Marchionini (1987) argues that people browse for three reasons: the search objective cannot be defined clearly, the cognitive burden is less than the structured formalized searching, or perhaps the retrieval system 'encourages' browsing. Hildreth (1987) notes that most users do not search for a known item, but rather browse because the process of exploration and discovery is essential to them. Serendipity is considered an important and valuable method of information gathering, which may or may not be a conscious strategy. As an information-seeking behaviour, Foster and Ford (2003) suggest that serendipity is not prominent in models proposed by researchers, including Kuhlthau's and Spink's (1997). Regardless of the variations in definitions, typologies, or reasons provided for browsing, many agree with Cove and Walsh (1988) that browsing may provide a "suitable paradigm" in an information system (p. 31).

Most researchers agree that children seem to prefer browsing for information in electronic environments rather than using planned search strategies (Large and Beheshti, 2000; Large, Beheshti and Moukdad, 1999; Schacter, Chung and Dorr, 1998). Children, and to some extent adults, have a "natural tendency to explore" (Borgman et al., 1995, p. 665), and are constantly browsing to discover the unknown (Chang, 1995). Borgman (1996, p. 495) states that users need "three layers of knowledge" to conduct effective searches: conceptual knowledge allows the user to formulate a query, semantic knowledge shows the user how and when to utilize various search options, and syntactic knowledge deals with the actual execution of the query. Children (as well as many adults) too often lack such knowledge. These layers of knowledge are not required in browsing, which can be used for a range of information-seeking activities, from well-defined tasks (known titles and authors) to semi-defined (specific subjects), and ill-defined (something of interest) tasks.

This suggests that systems that rely heavily upon browsing may prove both more attractive to children and more effective in meeting their information needs, at least during the exploratory/discovery stage of the ISP model. Browsing, however, may result in disorientation, particularly prevalent with hypertext-based systems as encountered on the Web. As users travel from node to node they may lose sight both of their information destination and their starting point.

Browsing relies upon recognition rather than recall; while in analytical searching users must recall from memory suitable search terms and assemble them into a coherent strategy, in browsing they are presented with several paths in the information space through which to navigate. In general, adults have better recognition memory than recall memory, but the difference between the two types of memory is more pronounced in children (Borgman et al., 1995). Lack of fully developed recall memory in children may result in their inability to form a schema, which they may need in formulating and revising search strategies (Borgman et al., 1995; Fidel et al., 1999; Nahl and Harada, 1996). However, children may also encounter browsing problems when they cannot easily match an information need to a browsing signpost. For example, Large et al. (2005) found that elementary school students chose a browsing approach when searching the Web so long as they could identify an appropriate entry point to a browsable site, but opted for keyword searching when the entry point was difficult to detect.

Browsing paths or options typically are visual and displayed on the screen from which users can choose. This emphasis on the visual (Cove and Walsh, 1988) has led some researchers to explore interface designs that rely heavily upon displayed images rather than displayed text to establish the path. For example, the Public Access Catalogue Extension (PACE) was designed as a novel visual interface for browsing library catalogues, to alleviate novice adult users' problems in coping with retrieval systems (Beheshti, 1992; Beheshti, Large, and Bialek, 1996). It was tested operationally in a college library against a second-generation online catalogue. More than 80 percent of students preferred PACE over the Library's traditional, text-based catalogue, and the simple browsing mechanism of PACE proved as effective in retrieval as the catalogue's Boolean-based system.

4.1. Metaphors for Browsing

Metaphors have been used in many systems to assist users to assimilate abstract concepts. A common example is the desktop metaphor (Erickson, 1990), which after several iterations is now commonly used in the vast majority of systems. PACE was based on the notion that a bibliographic record structure could be combined with better system design to provide the user with a mental 'picture' of the database. It proved successful because users readily and easily recognized the bookshelves and the books based on their experience in the libraries. They did not need to use an unfamiliar interface based on unknown mental models. As Bates (1989, p. 419) suggested:

> . . . if the interface can produce a picture on the screen that looks like the books on a shelf, the searcher can transfer a familiar experience to the automated system. So that if a mouse or similar device makes it possible to, in effect, move among the books, a familiar physical experience is reproduced and the searcher can take advantage of well-developed browsing skills.

The book metaphor has been used successfully in an online information system to assist users in search and navigation by utilizing their intuitive skills (Russell, 2001). Visual metaphors allow users to utilize "their cognitive, perceptual and intuitive skills to find data which may be of interest, but could be missed by search algorithms because they are not directly relevant to the query" (Young, 1996, *1. Introduction*). Metaphors, however, may also deter users if their expectations based on their mental models are not reasonably met by the system (cognitive dissonance); this can be avoided by utilizing familiar metaphors (Hamilton, 2000).

Children and young adults in particular tend to rely on visual information or visual cues rather than textual information (Fidel et al., 1999; Hirsh, 1999; Large and Beheshti, 2000). According to Nielsen (2002, paragraph 5) children like geographical navigational metaphors such as pictures of rooms, villages, and "other simulated environments that serve as an overview and entry point to various site or subsite features."

Borgman et al. (1991) in their first experiment with the Science Library project stated that their main objective in developing the Science Library Catalog was to "understand children's information-searching behavior sufficiently to design a system with powerful searching mechanisms that build on children's natural tendencies to explore, that can be used without prior training, and is within their range of skills and knowledge, while avoiding the need for typing skills, correct spelling, vocabulary knowledge, alphabetizing, and Boolean logic." (Borgman et al., 1995, p. 681) The Project, designed for elementary schoolchildren aged 9 to 12, provided access to a limited number of bibliographic records for science and technology monographs using a "bookshelf metaphor to correspond to children's mental models of a library catalog." (Hirsh, 1997, p. 603) The experiment showed that children had success rates averaging 80 percent with the Science Library Catalog. It also demonstrated that children do not like using tools such as catalogues and prefer to browse directly at the shelf, and that they find familiar metaphors highly appealing.

Druin and her colleagues (2003) make use of images in their experimental interface to the Children's International Digital Library (http://www.icdlbooks.org/) to facilitate navigation. Savage-Knepshield and Belkin (1999) predict that information visualization will be very much a part of the next IR wave, and will dramatically impact children's IR.

One specific application of visual interface design techniques is the virtual or three-dimensional interface. In such a system the user is immersed in the interface environment; he or she can explore that environment by moving around, examining or collecting objects, and so on. The best current examples of such virtual interfaces are to be found in the world of video games. Some development work has been undertaken with virtual interfaces to IR systems. For example, Christoffel and Schmitt (2002), Cubaud, Thiria, and Topol (1998), Das Neves, and Fox (2000), Kamiya et al. (2001) and Rauber and Merkl (2000), all have employed a library metaphor in designing three-dimensional interfaces.

4.2. Children's Views

Hanna, Risden, and Alexander (1997) are critical of researchers who assume that they can evaluate the usability of a product for children better than the children themselves. They believe that important benefits can be realized by collecting information directly from children as users of systems and no "amount of adult evaluation will find all the issues that children will stumble into as a matter of course" (p. 14). They argue that children in the 11 to 14 year age group are very easy to involve in usability testing; most will be comfortable with computers and with unfamiliar adults, and will enjoy carrying out activities.

In one of her recent works, Druin (2002) employed intergenerational teams to design and create children's information technologies. As part of her cooperative inquiry process she uses a combination of techniques from various participative design methodologies to partner professional designers with child users. She has employed cooperative inquiry to design and construct the International Children's Digital Library.

Bilal has investigated how grade-seven students in middle school might design Web search engine interfaces (Bilal, 2002a, 2003). She describes her approach as participatory design, with the participants as full design partners, but unlike in Druin's case, Bilal does not directly involve adults in the design process. She studied 11 grade-seven children working individually rather than as a design team and interviewed the students to discover their rationale for adding particular features to an IR interface. Bilal (2002a) concludes that children for most part are able to design the layout of an interface, outline the features they need, and describe the look and the appearance of these features. She strongly recommends that designers of children's search engines not only ensure that their engines are "cool" but also that they offer high usability.

In our investigations on designing children's Web portals, we began with focus group studies of existing portals specifically designed for children (Large, Beheshti, and Rahman, 2002a, 2002b). The focus groups consisted of 11 and 12 year old children, who provided us with feedback on the content, the interface, and the retrieval systems of four portals. We then embarked on a project to design two Web portals for two groups of children through intergenerational design teams (Large et al., 2003a, 2003b, 2004, 2005). One group consisted of eight grade six students, the other included six younger children in grade three.

In these studies, we utilized "Bonded Design", a new methodology to draw upon children's expertise at being children. The methodology is situated between Cooperative Inquiry (Druin, 1999) and Informant Design (Scaife et al., 1997) in terms of the degree of user involvement. In the Bonded Design methodology, as in Cooperative Inquiry, we believe in the ability of children to work as partners in all aspects of the design process. In employing Bonded Design, however, we also believe that children can participate in the design process in so far as their level of knowledge and expertise allows them. Above all, the child participates as expert at being children. In this respect, Bonded Design is similar to Informant Design.

The two intergenerational teams in our studies worked over a number of sessions starting with simple drawings and working toward a low-tech prototype (Large et al., 2003a, 2003b). Children in general suggested the following features for their portals:

- Several approaches to IR to suit individual needs and preferences: a keyword search box and a natural language search box, a broad topic menu consisting of broad subject headings, and an alphabetical browser, and a scrollable timeline (in the context of a portal on Canadian history)
- Links to other Web portals and search engines, specifically to Google, MSN and Yahoo
- Advanced keyword search features that would allow Boolean term combination and phrase searching—even though research has demonstrated that children have difficulties in using Boolean logic (Borgman, 1996; Nahl and Harada, 1996; Schwartz, 1998; Solomon, 1993).
- Options for restricting searches to visual and audio contents
- Email and chat rooms as interactivity tools in an educational environment (even though children were familiar with the potential dangers)
- Educational games, puzzles and quizzes (even though they were recognized as having a distraction component)
- Within a Canadian context, bilingual interfaces
- In-context help facilities
- Personalization features

The low-tech portals designed by the two intergenerational teams were then converted into operational Web portals and tested by 32 students in grades three and six (Large et al., in press). The results show that the children enthusiastically endorsed the portals. Many features that were suggested by the Bonded Design team members were in fact embraced by the children in the test groups, demonstrating the ability of the youngsters to participate equally in the design process, and more important, showing their valuable contribution to the design of functional Web portals for young users.

Through one of the many sessions in the Bonded Design process, we engaged the older children in grade six in discussions about new systems for retrieving information for classroom projects and educational purposes. The session was entirely devoted to discussions about a novel virtual reality interface designed primarily rather like a computer game. Based on earlier research on the use of familiar metaphors, browsing, and visualization, we were primarily interested in provoking the students to engage in debate and brainstorming on a different and non-conventional approach to designing a portal for children. In addition to the conventional portal which the team had developed earlier, children suggested a more radical virtual reality design where users would move through a first-person virtual environment browsing for information. In the VR system, rather as in a computer game, users would explore the Web by moving through virtual space. All but one student were extremely enthusiastic about such an approach. All agreed, however, that in order to meet the needs and requirements of all users the VR environment should complement rather than replace any conventional portal interfaces.

In designing their individual VR portals, all students (with one exception) relied heavily on the use of metaphors. The metaphors are based on familiar representations of the real world (Beheshti et al., 2004):

- A calendar displayed prominently in one design shows the days and months of the year representing "Current Events"
- Buildings and roads that represent different search options, interactivity, and guidelines for navigation
- Historical buildings representing time lines
- Museums that guide users to simulated artifacts
- Playgrounds that direct users to "edutainment"
- Avatars that either symbolize the user or an "expert" that can help the user.

5. FUTURE DIRECTIONS: AN ALTERNATIVE IR SYSTEM FOR CHILDREN

Fidel et al. (1999) suggest that summarizing the findings of various studies on children's information seeking behavior may prove difficult since these studies do not all use the same methodologies, age groups, or similar demographics. Nevertheless, we can make some general observations, which may be used in developing a new IR system for children. The Zone of Proximal Development argues for a system that can implement supportive context. In fact, any retrieval system for young people would be well advised to incorporate tools and mechanisms that encourage them to continue using the system.

The notion of a Search Pal that provides context-sensitive help whilst displaying appropriate emotions would be an ideal way to support the children in their information-seeking situations. Children in the Bonded Design teams suggested a similar notion in the context of the Web portals that they were helping to develop. They would have liked to incorporate context-sensitive help facilities within the portal to assist users in retrieving the required information. The idea of including emotions within such a system did not occur to them. Two factors may have confined their imagination: they had never encountered such a system in their limited experience; and the adult members of the Bonded Design Team did not point out such a novel idea.

We also know that IR systems should reduce the anxiety levels of children by helping them cope with information overload, and feel in control of the retrieval environment. They need to feel positive about the task. Ease of use, mobility and navigation, and "fun" are the main factors that contribute to reducing children's anxiety levels. Encouragement, enthusiasm, exploration and discovery are traits that can be built into empathetic IR systems. Such a system can include hints, clues, prizes, and other motivational tools, which are activated through intelligent feedback mechanisms to encourage children in their quest for information.

Systems designed for children must be engaging. Engagement may be indicative of the likes and dislikes of children and their level of satisfaction, which can be measured by observing their behaviour (Hanna, Risden, and Alexander, 1997). Hanna et al. (1999) believe that usability research has identified ease of use as a critical factor in engagement, "and as such is key to every child's product if it is to be a success" (p. 4). Although there are many definitions of usability, most experts would agree that it refers to the ability of the user to easily and intuitively understand and navigate the components of the interface (Head, 1997; Nielsen, 2000; Rose, Shneiderman, and Pleasant, 1995; Rubin, 1994). Usability is directly related to the cognitive and affective characteristics of children. Operationally it may be defined as visual aesthetics as judged by children, conceptual and linguistic clarity, compatibility with the task, comprehensibility, consistency, controllability, familiarity, flexibility, predictability, simplicity, and responsiveness (Galitz, 2002).

The traditional usability approach to system design, however, may tend to view users in a limited fashion. Emotions, hopes, fears, and joy have not for most part been considered as standard variables to be studied in conventional usability research. Recently, the importance of including these factors in the human-computer interaction research has been discussed in the literature (Blythe, 2004). Particularly significant in systems designed for children is the 'fun' factor. In relation to developing a museum site for young users, Rizzo et al. (2003, p. 198) state that the interface "should be equally attractive, fun, long lasting, and yet offer sound pedagogical learning opportunities." Indeed, "children are strong in their declaration that they expect to have fun using technology" (Shneiderman 2004, p. 49) and they often link the idea of fun to challenges, social interaction, and control over their world (Druin and Inkpen, 2001).

Bilal, in a series of articles (2000, 2001, 2002a) investigated how grade–seven students interacted with a children's Web portal, Yahooligans! In three task-based studies she asked children to find information using the portal. She offers a number of suggestions to portal designers based on her findings: they should "develop search engines with powerful searching and browsing mechanisms that built on children's cognitive and physical behaviors to search, browse, navigate and explore information" (Bilal, 2000, p. 662). She proposes more search and browse instructions accompanied with examples, as well as context-sensitive help and online tutorials, a natural-language interface, output ranking, simple screen displays, and spell checking.

Failure to consider engagement as an important factor in designing an IR system may result in little interaction between the child and the system, and result in overall dissatisfaction with the system. Children passively viewing rather than actively interacting with the information system quickly become bored (Said 2004). Among a list of five features contributing to the experience of engagement, Said (2004, p. 169) mentions "simulation interaction which allows the child to act on behalf of others i.e. to role-play." In the initial *state of uncertainty* in the ISP model, motivation is a particularly important and crucial factor (Kuhlthau, 1999) in encouraging the child to continue using the information system. Through engagement and motivation, the child would feel empowerment and in control of the system.

We know that browsing is an alternative search strategy used by children, from seven year olds (Cooper, 2002) to 18 year olds (Fidel et al., 1999), requiring significantly less cognitive load than analytical searching. Browsing is a visual activity. To accommodate browsing, familiar metaphors may be utilized in designing IR systems. Familiar metaphors may be best presented in three dimensional virtual reality systems to increase their value. Virtual environments are highly effective means of teaching and transferring knowledge, particularly historical information, and can have a strong motivational impact (Bricken, 1991). These systems provide opportunities to experience environments which, for reasons of time, distance, scale, and safety, would not otherwise be available to many children, especially those with disabilities (Standen, Brown, and Cromby, 2001). Early exposure to virtual environments may both facilitate the efficiency and capacity of children's learning and provide advance organizers for later learning experiences (Dede, 1998).

The younger Bonded Design team members used avatars in their 3D drawings, to include other characters in their futuristic worlds. In an ideal IR system, avatars may be used to depict intelligent agents providing supportive context, which also display appropriate emotions. Avatars have been used in games for the past decade or so, and more recently research has been conducted on incorporating emotions in the avatars for the benefit of children (Antle, 2004; Roussou, 2004).

An alternative IR system for children will therefore include a virtual reality environment based on a familiar metaphor designed for browsing, and avatars that are intelligent enough to provide context sensitive help and tutoring whilst displaying appropriate emotions. The virtual environment will include clues, treasures, and other fun-filled tools to engage the children and keep them engaged during the course of information seeking. As our young colleagues in the Bonded Design team suggested such a system should be integrated with a conventional portal or search engine, to be used as an alternative tool for those children who like to use browsing rather than analytical strategies.

REFERENCES

Agosto, D. (2002). A Model of Young People's Decision-Making in Using the Web. *Library and Information Science Research*, 24(4), 311–341.

Antle, A. (2004). Supporting Children's Emotional Expression and Exploration in Online Environments. In *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community*. *Interaction Design and Children*. New York: ACM Press.

Apted, S. M. (1971). General Purposive Browsing. *Library Association Record*, 73, 228–30.

Bandura, A. and Huston, A. C. (1963). Identification as a Process of Incidental Learning. *Journal of Abnormal Social Psychology,* 63, 311–318.

Bates, M. J. (1989). The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review,* 13(5), 407–424.

Baylor, A. (1999). Intelligent Agents as Cognitive Tools for Education. *Educational Technology*, March/April, 39(2), 36–40.

Beheshti, J. (1992). Browsing Through Public Access Catalogs. *Information Technology and Libraries*, 11(3), 220–228.

Beheshti, J., Large, V., and Bialek, M. (1996). Public Access Catalogue Extension (PACE): A Browsable Graphical Interface. *Information Technology and Libraries,* 15(4), 231–240.

Beheshti, J., Large, A., Nesset, V., and Bowler, L. (2004). Children's Web Portals: Novel Designs. In *Proceedings of the International Conference on Information and Knowledge Engineering (IKE '04)* (pp. 399–405). Las Vegas: CSREA Press.

Belkin, N. J. (1980). Anomalous States of Knowledge as a Basis for Information Retrieval. *The Canadian Journal of Information Science*, 5, 133–143.

Bilal, D. (2000). Children's Use of the Yahooligans! Web Search Engine: 1. Cognitive, Physical and Affective Behaviors on Fact-Based Search Tasks. *Journal of the American Society for Information Science*, 51(7), 646–665.

Bilal, D. (2001). Children's Use of the Yahooligans! Web Search Engine. II. Cognitive and Physical Behaviors on Research Tasks. *Journal of the American Society of Information Science and Technology*, 52(2), 118–136.

Bilal, D. (2002a). Children Design Their Interfaces for Web Search Engines: A Participatory Approach. In *Proceedings of the 30th Annual Conference of the Canadian Association for Information Science* (pp. 204–214). In: L. Howarth, C. Cronin and A. Slawek (Eds.). CAIS .

Bilal, D. (2002b). Children's Use of the Yahooligans! Web Search Engine. III. Cognitive and Physical Behaviors on Fully Self-Generated Search Tasks. *Journal of the American Society of Information Science and Technology*, 53(13), 1170–1183.

Bilal, D. (2002c). Perspectives on Children's Navigation of the World Wide Web: Does the Type of Search Task Make a Difference? *Online Information Review*, 26(2), 108–117.

Bilal, D. (2003). Draw and Tell: Children as Designers of Web Interfaces. In *Proceedings of the 66th Annual Meeting of the American Society for Information Science and Technology: Humanizing Information Technology: From Ideas to Bits and Back* (pp. 135–141). Medford, NJ: Information Today Inc.

Bilal, D. and Kirby, J. (2002). Differences and Similarities in Information Seeking: Children and Adults as Web Users. *Information Processing and Management*, 38(5), 649–670.

Blythe, M. (2004). Interview with Patrick Jordan. *Interactions*, 11(5), 40–41.

Blumenfeld, P. C., Solway, E., Marx, R., Krajcik, J. S., Guzdial, M., and Palinscar, A. (1991). Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning. *Educational Psychologist*, 26(3/4), 369–398.

Borgman, C. (1996). Why are Online Catalogs Still Hard to Use? *Journal of the American Society for Information Science,* 47(7), 493–503.

Borgman, C. L., Gallagher, A. L., Walter, V. A., and Rosenberg, J. B. (1991). The Science Library Catalog Project: Comparison of Children's Searching Behavior in a Direct Manipulation and a Keyword Search System. In: J. M. Griffiths (Ed.), In *Proceeding of the 54th Annual Meeting of the American Society for Information Science*, 28 (pp. 162–169). Medford, NJ: Learned Information, Inc.

Borgman, C. L., Hirsh, S. G., Walter, V. A., and Gallagher, A. L. (1995). Children's Searching Behavior in Browsing and Keyword Searching Online Catalogs: The Science Library Catalog Project. *Journal of the American Society for Information Science*, 46(9), 663–684.

Bowler, L., Large, A., and Rejskind, G. (2001). Primary School Students' Information literacy and the Web. *Education for Information*, 19, 201–223.

Bricken, M. (1991). *No Interface to Design. Cyberspace: The First Steps*. MIT Press, Cambridge, MA.

Brin, S. and Page, L. (1997). The Anatomy of a large-scale hypertextual Web search engine. http://www-db.stanford.edu/~backrub/google.html [Accessed March 31, 2005].

Broch, E. (2000). Children's Search Engines From an Information Search Process Perspective. *School Library Media Research.* Available at http://www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume32000/childrens.htm [Accessed September 28, 2003].

Brown, A. L., Ellery, S., and Campione, J. C. (1998). Creating Zones of Proximal Development Electronically. In: J. G. Greeno and S. Goldman (Eds.), *Thinking Practices in Mathematics and Science Learning* (pp. 341–368). Hillsdale, NJ: Erlbaum.

Chang, S. (1995). *Towards a Multidimensional Framework for Understanding Browsing*. Unpublished doctoral dissertation, Rutgers, The State University, New Jersey.

Chang, S. J. and Rice, R. E. (1993). Browsing: A Multidimensional Framework. In: M. E. Williams (Ed.), *Annual Review of Information Science and Technology*, 28 (pp. 231– 276), Learned Information: Medford, New Jersey Christoffel, M. and Schmitt, B. (2002). Accessing Libraries as Easy as a Game. *Second International Workshop on Visual Interfaces to Digital Libraries, July 18, 2002, Portland, Oregon*. Available at http://vw.indiana.edu/visual02/slides-christoffel.pdf [Accessed March 31, 2005].

Clarebout, G., Elen, J., Johnson, W. L., and Shaw, E. (2002). Animated Pedagogical Agents: An Opportunity to be Grasped? *Journal of Educational Multimedia and Hypermedia*, 11(3), 267–286.

Cole, C. (1999). Activity of Understanding a Problem During Interaction with an "Enabling" IR System: Modeling Information Flow. *Journal of the American Society for Information Science*, 50(6), 544–552.

Cole, C. (2000). Interaction with an Enabling Information Retrieval System: Modeling the User's Decoding and Encoding Operations. *Journal of the American Society for Information Science*, 51(5), 417–426.

Cooper, L. Z. (2002). A Case Study of Information-Seeking Behavior in 7-year-Old Children in a Semi-Structured Situation. *Journal of the American Society for Information Science and Technology*, 53(11), 904–922.

Cove, J. F. and Walsh, B. C. (1988). Online Text Retrieval Via Browsing. *Information Processing and Management*, 24(1), 31–37.

Cubaud, P., Thiria, C., and Topol, A. (1998). Experimenting a 3D Interface for the Access to a Digital Library. In *Proceedings of the Third ACM Conference on Digital Libraries*. Pittsburgh, PA.

Das Neves, F. A. and Fox, E. A. (2000). A Study of Users Behavior in an Immersive Virtual Environment for Digital Libraries. In *Proceedings of the Fifth ACM Conference on Digital Libraries* (pp. 103–112), San Antonio, TX.

Dede, C. (1998). The Scaling-Up Process for Technology-Based Educational Innovations. In: *Learning with Technology* (1998 ASCD Yearbook), edited by the author. Alexandria, VA: Association for Supervision and Curriculum Development.

Dervin, B. (1999). Chaos, Order, and Sense-Making: A Proposed Theory for Information Design. In: R. Jacobson (Ed.), *Information Design* (pp. 37–57). Cambridge, Mass.: The MIT Press..

Druin, A. (1999). Cooperative Inquiry: Developing New Technologies for Children with Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* M. Williams and M. Altom. (chairpersons) (pp. 592–599). New York, NY: ACM Press.

Druin, A. and Inkpen, K. (2001) When are Personal Technologies for Children? *Personal and Ubiquitous Computing,* 5(3), 191–194.

Druin, A. (2002). The Role of Children in the Design of New Technology. *Behaviour and Information Technology*, 21(1), 1–25.

Druin, A., Bederson, B. B., Weeks, A., Farber, A., Grosjean, J., Guha, M. L., Hourcade, J. P., Lee, J., Liao, S., Reuter, K., Rose, A., Takayama, Y., and Zhang, L. (2003). The International Children's Digital Library: Description and Analysis of First Use. *First Monday*, 8(5). Available at http://www.firstmonday.dk/issues/issue8_5/druin/index.html [Accessed March 31, 2005].

Erickson, T. (1990). Working with Interface Metaphors. In: *The Art of Human- Computer Interface Design*. Edited by Brenda Laurel. New York: Addison-Wesley.

Fidel, R., Davies, R., Douglass, M., Holder, J., Hopkins, C., Kushner, B., Miyagishima, B., and Toney C. (1999). A Visit to the Information Mall: Web Searching Behavior of High School Students. *Journal of the American Society for Information Science*, 50, 24–37.

Foster, A. and Ford, N. (2003). Serendipity and Information Seeking: An Empirical Study. *Journal of Documentation*, 59(3), 321–340.

Galitz, W. O. (2002). *The Essential Guide to User Interface Design*. 2nd Ed. New York: Wiley.

Hamilton, A. (2000). Metaphor in Theory and Practice: The Influence of Metaphors on Expectations. *ACM Journal of Computer Documentation*, 24(4), 237–253.

Hanna, L., Risden, K., and Alexander, K. (1997). Guidelines for Usability Testing with Children. *Interactions*, 4(5), 9–14.

Hanna, L., Risden, K., Czerwinski, M., and Alexander, K. (1999). The Role of Usability Research in Designing Children's Computer products. In: A. Druin (Ed.). *The Design of Children's Technology* (pp. 3–26). San Francisco, CA: Morgan Kaufmann Publishers.

Head, A. J. (1997). Web Usability and Essential Interface Design Issues. In: M. E. Williams (Ed.), In *Proceedings of the 18th National Online Meeting; New York, NY* (pp. 157–163). Information Today: Medford, NJ: .

Herner, S. (1970). Browsing. *Encyclopedia of Library and Information Science*, 3, 408–415.

Hidi, S. (1990). Interest and Its Contribution as a Mental Resource for Learning. *Review of Educational Research*, 60, 549–571.

Hildreth, C.R. (1987). Beyond Boolean: Designing the Next Generation of Online Catalogs. *Library Trends*, 35(4), 647–667.

Hirsh, S. G. (1997). How Do Children Find Information on Different Types of Tasks? Children's Use of the Science Library Catalog, *Library Trends*, 45, 602–622.

Hirsh, S. G. (1999). Children's Relevance Criteria and Information Seeking on Electronic Resources. *Journal of the American Society for Information Science*, 50(14), 1265–1283.

Hyman, R. J. (1971). Access to Library Collections: Summary of a Documentary and Opinion Survey on the Direct Shelf Approach and Browsing. *Library Resources and Technical Services*, 15, 479–491.

Julien, H., McKechnie, L., and Hart, S. (2004). A Content Analysis of Affective Issues in Library and Information Science Systems Work, [Summary of a Research Note Delivered at the ISIC 2004 conference, Dublin, 1–3 September, 2004] *Information Research*, 10(1) summary 6. Available at http://informationr.net/ir/10-1/abs6.html. [Accessed March 31, 2005].

Kafai, Y. and Bates, M. (1997). Internet Web-Searching Instruction in the Elementary Classroom: Building a Foundation for Information Literacy. *School Library Media Quarterly,* 25(2), 103–111.

Kamiya, T., Lu, S., Hara, M., and Miyai, H. (2001) Development of Electronic Library Interface with 3D Walk-Through and CG Librarian. *Technical Report of Information Processing Society of Japan,* IM19-5, pp. 27–34. (in Japanese). [English available at http://www.ipsj.or.jp/members/SIGNotes/Eng/33/1994/019/article005.html, accessed Aug., 2003]

Kuhlthau, C. C. (1991). Inside the Search Process: Information Seeking from the User's Perspective. *Journal of the American Society for Information Science*, 42(5), 361–371.

Kuhlthau, C. C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. Westport, CT: Ablex Publishing.

Kuhlthau, C. C. (1997). Learning in Digital Libraries: An ISP Approach. *Library Trends*, 45(4), 708–724.

Kuhlthau, C. C. (1999). Accommodating the User's ISP: Challenges for Information Retrieval Systems Designers. *Bulletin of the American Society for Information Science,* 25 (3), 1–7. Available at http://www.asis.org/Bulletin/Feb-99/kuhlthau.html [Accessed March 31, 2005]

Kuhlthau, C. C. (2004). *Seeking Meaning: A Process Approach to Library and Information Services*. (2nd Ed.). Westport, CT: Libraries Unlimited.

Kuhlthau, C. C. (2005). Towards Collaboration between Information Seeking and Information Retrieval. *Information Research*, 10(2) paper 226. Available at http://InformationR.net/ir/10-2/paper226.html. [Accessed March 31, 2005].

Lancaster, F.W. (1968). *Information Retrieval Systems: Characteristics, Testing, and Evaluation*. New York: Wiley, pp. 181–182.

Large, A. (2005). Children, Teenagers, and the Web. *Annual Review of Information Science and Technology* (pp. 347–392). Washington, DC: American Society for Information Science.

Large, A. and Beheshti, J. (2000). The Web as a Classroom Resource: Reactions from the Users. *Journal of the American Society for Information Science*, 51(12), 1069–1080.

Large, A., Beheshti, J., and Breuleux, A. (1998). Information Seeking in a Multimedia Environment by Primary School Students. *Library and Information Science Research*, 20(4): 343–376.

Large, A., Beheshti, J., and Moukdad, H. (1999). Information Seeking on the Web: Navigational Skills of Grade-Six Primary School Students. *Knowledge: Creation, Organization and Use: Proceedings of the 62nd ASIS Annual Meeting* (pp. 84–97), Information Today: Medford, NJ, Inc.

Large, A., Beheshti, J., and Rahman, T. (2002a). Gender Differences in Collaborative Web Searching Behavior: An Elementary School Study. *Information Processing and Management*, 38, 427–443.

Large, A., Beheshti, J., and Rahman, T. (2002b). Design Criteria for Children's Web Portals: The Users Speak Out. *Journal of the American Society for Information Science and Technology,* 53(2), 79–94.

Large, A., Beheshti, J., Nesset, V., and Bowler, L. (2003a). Children as Designers of Web Portals. In *Proceedings of the 66th Annual Meeting of the American Society for Information Science and Technology: Humanizing Information Technology: From Ideas to Bits and Back* (pp. 142–149). Medford, NJ: Information Today, Inc.

Large, A., Beheshti, J., Nesset, V., and Bowler, L. (2003b). Children as Web Portal Designers: Where Do We Start? In: W. C. Peekhaus and L. F. Spiteri (Eds.), *Bridging the Digital Divide: Equalizing Access to Information and Communication Technologies.* In *Proceedings of the 31st Annual Conference of the Canadian Association for Information Science* (pp. 139–152), Dalhousie University, May 30–June 1, 2003. Halifax: CAIS.

Large, A., Beheshti, J., Nesset, V., and Bowler, L. (2004). Designing Web Portals in Intergenerational Teams: Two Prototype Portals for Elementary School Students. *Journal of the American Society for Information Science and Technology*, 55(13), 1150–1154.

Large, A., Beheshti, J., Nesset, V., and Bowler, L. (2005). Web Portal Characteristics: Children as Designers and Evaluators. *Data, Information and Knowledge in a Networked World:* In *Proceedings of the 33rd Annual Conference of the Canadian Association for Information Science,* University of Western Ontario, June 2–June 4, 2005. London, ON: CAIS.

Large, A., Nesset, V., Beheshti, J., and Bowler, L. (in press). Design Criteria for Children's Web Portals: A Comparison of Two Studies. *Canadian Journal of Information and Library Science*.

Leide, J. E., Large, A., Beheshti, J., Martin, B., and Cole, C. (2003). Visualization Schemes for Domain Novices Exploring a Topic Space: The Navigation Classification Scheme. *Information Processing and Management*, 39(6), 923–940.

Lepper, M. (1988). Motivational Considerations in the Study of instruction. *Cognition and Instruction,* 5(4), 289–309.

Lepper, M. and Chabay, R. (1988). Socializing the Intelligent Tutor: Bringing Empathy to Computer Tutors. In: H. Mandl and A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems.* New York: Springer Verlag.

Marchionini, G. (1987). An Invitation to Browse: Designing Full-Text Systems for Novice Users. *Canadian Journal of Information Science*, 12(3/4), 69–79.

Marchionini, G. (1995). *Information Seeking in Electronic Environments*. New York: Cambridge University Press.

Mayer, R. E. (1997). Learners as Information Processors: Legacies and Limitations of Educational Psychologies Second Metaphor. *Educational Psychologist*, 31(3/4), 151–161.

Moreno, R., Mayer, R., Spires, H., and Lester, J. (2001). The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact with Animated Pedagogical Agents? *Cognition and Instruction*, 19(2), 177–213.

Moundridou, M. and Virvou, M. (2002). Evaluating the Personal Effect of an Interface Agent in a Tutoring System. *Journal of Computer Assisted Learning*, 18, 253–261.

Nahl, D. (1995). The User-Centered Revolution: 1970–1995. In: A. Kent (Ed.) *Encyclopedia of Library and Information Science*, 62(Suppl. 25), (pp. 313–371). New York: Marcel Dekker.

Nahl, D. and Harada, V. H. (1996). Composing Boolean Search Statements: Self-Confidence, Concept Analysis, Search Logic and Errors. *School Library Media Quarterly*, 24, 199–207.

Negroponte, N. (1997). Agents: From Direct Manipulation to Delegation. In: J. M. Bradshaw (Ed.). Software Agents (pp. 57–66). Menlo Park, CA: MIT Press.

Nielsen, J. (2000). *Designing Web Usability*. Indianapolis, Ind.: New Riders.

Nielsen, J. (2002). Kid's Corner: Website usability for children. *Useit.com Alertbox*. Available at http://www.useit.com/alertbox/20020414.html. [Accessed March 31, 2005].

Neuman, D. (1995). High school students' use of databases: Results of a national Delphi study. *Journal of the American Society for Information Science,* 46(4), 284–298.

Pejtersen, A. M. (1989). A Library System for Information Retrieval Based on a Cognitive Task Analysis and Support By an Icon-Based Interface. In: *ACM SIGIR Conference* In *Proceedings, Boston, 1989*. New York: Association for computing Machinery.

Perez, E. (2002). Intelligent Agents: It's Nice to Get Stuff Done for You. *Online.* May/June, 51–56.

Piaget, J. and Inhelder, B. (1969). *The Psychology of the Child*. New York: Basic Books.

Picard, R. W. (2003). Affective Computing: Challenges. *International Journal of Human-Computer Studies*, 59(1–2), 55–64.

Rauber, A. and Merkl, D. (2000). SOMLib: A Digital Library System Based on Neural Networks. Available at http://www.ifs.tuwien.ac.at/~andi/somlib/publications/pub_html/rau_ecdl99/rau_ecdl99.html. [Accessed March 31, 2005].

Rice, R. E., McCreadie, M., and Chang, Shan-Ju, L. (2001). *Accessing and Browsing Information and Communication*. Cambridge, Mass: The MIT Press.

Rizzo, A., Marti, P., Decortis, F., Rutgers, J., and Thursfield, P. (2003). Building Narrative Experiences for Children Through Real Time Media Manipulation: POGO World. In: M. A. Blythe, K. Overbeeke, A.F. Monk, and P. C. Wright (Eds.). *Funology: From Usability to Enjoyment*. Boston: Kluwer Academic Publishers.

Rose, A., Shneiderman, B., and Plaisant, C. (1995). An Applied Ethnographic Method for Redesigning User Interfaces. In *ACM Proceedings of DIS 95, Symposium on Designing interactive Systems: Processes, Practices, methods and Techniques* (pp. 115–122). Ann Arbor, MI.

Roussou, M. (2004). Learning By Doing and Learning Through Play: An Exploration of Interactivity in Virtual Environments for Children. *Computers in Entertainment*, 2(1), 10.

Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design and Conduct Effective Tests*. New York, NY: Wiley.

Russell, J. (2001). Book Metaphor: Friend or Foe? In *Proceedings of the 19th Annual International ACM Special Interest Group for Design of Communications Conference on Computer Documentation* (pp. 180–185). Sante Fe, New Mexico.

Safe and Smart: Research and Guidelines for Children's Use of the Internet. National School Boards Foundation. Available at http://www.nsbf.org/safe-smart/full-report.htm. [Accessed March 31, 2005].

Said, N. S. (2004). An Engaging Multimedia Design Model. In *Proceedings of IDC:* New York (pp. 169–172) NY: ACM Press.

Saracevic, T. (1999). Information Science. *Journal of the American Society for Information Science*, 50(12), 1051–1063.

Savage-Knepshield, P. A. and Belkin, N. J. (1999). Interaction in Information Retrieval: Trends Over Time. *Journal of the American Society for Information Science*, 50(12), 1067–1082.

Scaife, M., Rogers, Y., Aldrich, F., and Davies, M. (1997). Designing for or Designing With? Informant Design for Interactive Learning Environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 343–350). Pemberton, S. (Ed.), New York, NY: ACM PressSchacter, J., Chung, G. K., and Dorr, A. (1998). Children's Internet Searching on Complex Problems: Performance and Process Analyses. *Journal of the American Society for Information Science*, 49, 840–849.

Schwartz, C. (1998). Web Search Engines. *Journal of the American Society for Information Science*, 49(1), 973–982.

Siegler, R. S. (1998). *Emerging Minds: The Process of Change in Children's Thinking*. New York: Oxford University Press.

Shaw, E., Johnson, W. L., and Ganeshan, R. (1999). Pedagogical Agents on the Web. In *Proceedings of the Third International Conference on Autonomous Agents*. New York: ACM Press.

Shneiderman, B. (2004). Designing for Fun: How Can We Design User Interfaces to Be More Fun? *Interaction,* 11(5), pp. 48–50.

Solomon, P. (1993). Children's Information Retrieval Behavior: A Case Analysis of an OPAC. *Journal of the American Society for Information Science*, 44 (5), 245–264.

Standen, P. J., Brown, D. J., and Cromby, J. J. (2001). The Effective Use of Virtual Environments in the Education and Rehabilitation of Students with Intellectual Disabilities. *British Journal of Educational Technology*, 32(3), 289–299.

Tsai, C., Lin, S., and Tsai, M. (2001). Developing an Internet Attitude Scale for High School Students. *Computers and Education*, 37, 41–51.

Vandergrift, K. E. (1989). Are Children and Teenagers Second-Class Users? *Library Resources and Technical Services*, 33(4), 393–399.

Vise, D. (2003). AOL's Appeal to Youth Strategy to Hang on to Subscribers Targets Kids. *Washington Post*, September 10, p.E1.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wallace, R. M., Kupperman, J., Krajcik, J., and Soloway, E. (2000). Science on the Web: Students Online in a Sixth-Grade Classroom. *Journal of the Learning Sciences*, 9(1), 75–104.

Walter, V. A. (1994). The Information Needs of children. *Advances in Librarianship*, 18, 111–129.

Watson, J.S. (1998). "If You Don't Have It, You Can't Find It". A Close Look at Students' Perceptions of Using Technology. *Journal of the American Society for Information Science*, 49(11), 1024–1036.

Wilson, T. D. (1981). On User Studies and information needs. *Journal of Documentation,* 37(1), 3–15.

Wilson, T. D. (1997). Information Behavior: An Interdisciplinary Perspective. *Information Processing and Management*, 33(4), 551–572.

Young, Peter. (1996). Three Dimensional Information Visualisation. *Computer Science Technical Report, No. 12.96*, University of Durham, Department of Computer Science, November. Available at http://vrg.dur.ac.uk/misc/PeterYoung/pages/work/documents/lit-survey/IV-Survey/[Accessed March 31, 2005].

SECTION IV

CIR TECHNIQUES

CHAPTER 9

IMPLICIT FEEDBACK: USING BEHAVIOR TO INFER RELEVANCE

DIANE KELLY
University of North Carolina at Chapel Hill

1. INTRODUCTION

The concept of relevance has a rich history in information retrieval (IR) that dates back well over 40 years (Borlund, 2003) and is necessarily a part of any theory of information-seeking and retrieval. Relevance feedback also has a long history in IR (Salton, 1971) and is considered an important part of interactive IR (Spink and Losee, 1996). Relevance feedback techniques often require users to explicitly provide feedback to the system, by, for instance, specifying keywords, selecting, evaluating and marking documents, or answering questions about their interests. The feedback that users provide during these interactions has been used for a variety of IR techniques and applications including query expansion, term disambiguation, user profiling, filtering and personalization.

Empirical studies have led to the general finding that users of interactive IR systems desire explicit relevance feedback features and, in particular, term suggestion features (Beaulieu, 1997; Belkin et al., 2001; Koenemann and Belkin, 1996). However, much of the evidence from laboratory studies has indicated that relevance feedback features are not used. While users often report a desire for relevance feedback and term suggestion, they do not actually use these features during their searching activities. Several reasons can be given for why this disparately exists. Users may not have additional cognitive resources available to operate the relevance feedback feature. While the extra effort required to operate the feature may seem trivial, the user is already potentially involved in a complex and cognitively burdensome task. Increased effort would be required for both learning the new system and operating its features. When features require more effort and additional cognitive processing than they appear to be worth, they may be abandoned all together. Furthermore, if relevance feedback features are not implemented as part of the routine search activity, they may be forgotten, no matter how helpful they are. This research, in part, has lead to the general belief that users are unwilling to engage in explicit relevance feedback.

Recently (Anick, 2003) demonstrated in a web-based study, that users made use of a term suggestion feature to expand and refine their queries, thus things may be changing. These results suggest the potential of term suggestion features in some types of information-seeking environments, especially for single session interactions. Hence it may just be the case that traditional relevance feedback interfaces have not effectively elicited feedback from users or optimally integrated relevance feedback features into current information interaction models.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 169–186.

However, there are some problems with the sustainability of using explicit relevance feedback to personalize information-seeking interactions and use over time. These techniques typically require users to specify keywords or spend time selecting and marking documents. When considered from a long-term perspective, it seems unrealistic to expect users to read and rate a large number of documents on a daily basis, if they are even willing to do so. Because it is likely that these activities will appear to the user as irrelevant rather than necessary, it is unlikely that users will engage in explicit feedback activities. If users do not engage in relevance feedback activities, then there is no evidence upon which to provide personalization. Implicit relevance feedback techniques offer some solution to this problem.

Implicit relevance feedback techniques unobtrusively obtain information about users, their needs and document preferences, by watching their natural interactions with systems. Implicit feedback techniques are attractive methods for tailoring system responses to individual user preferences since they do not require additional effort on the part of the user. Some behaviors that have been most extensively investigated as sources of implicit feedback include reading time, saving, printing, selecting and referencing. The primary advantage to using implicit feedback techniques is that such techniques remove the cost to the user of providing explicit feedback. Implicit measures are generally thought to be less accurate than explicit measures (Nichols, 1997), but since large quantities of implicit data can be gathered at no extra cost to the user, they are thought to be attractive alternatives to explicit techniques.

The purpose of this chapter is to present a framework for understanding and studying how behavior can be used as implicit relevance feedback. This chapter begins with a classification and discussion of behaviors that have been used as implicit relevance feedback. This is followed by a general discussion and characterization of implicit feedback research. A selection of this research is then presented to illustrate how such studies have been conducted and how feedback has typically been measured and used. This chapter ends with a discussion of key issues and problems associated with implicit feedback and challenges for future research.

2. CLASSIFICATION OF BEHAVIORS USED FOR IMPLICIT FEEDBACK

Nichols (1997) provided the first classification of potential behavioral sources of implicit feedback by categorizing the actions that a user might be observed performing during information-seeking. For instance, reading time is considered as an action that allows one to "Examine" an information object. Nichols (1997) goes on to discuss the potential of, and the problems with, various sources of implicit feedback, as well as the costs. Specifically, Nichols (1997) states, "each implicit rating will probably contain less value than an explicit rating but the appropriate cost-benefit trade-off for different types of implicit data will have to be determined empirically." Oard and Kim (2001), Claypool et al. (2001), and Kelly and Teevan (2003) followed-up this initial classification with conceptual classifications of their own. Building on Nichols's (1997) classification, Oard and Kim (2001) proposed a framework of observable behaviors, which could be used to model interests using two axes, *Behavior Category* and *Minimum Scope*.

Confidence in Evidence	Behavior Category	Minimum Scope: Segment	Minimum Scope: Object	Minimum Scope: Class
		Precision: More		Less
Weak	**Examine**	View Listen Scroll Find Query	Select	Browse
	Retain	Print	Bookmark Save Delete Purchase Email	Subscribe
	Reference	Copy-and-paste Quote	Forward Reply Link Cite	
	Annotate	Mark up	Rate Publish	Organize
Strong	**Create**	Type Edit	Author	

Figure 1: Classification of behaviors for implicit feedback

This classification is displayed in Figure 1.

It should be noted that the classification proposed by Oard and Kim (2001) includes one behavior that is typically used as explicit feedback, rate. Rate is the traditional method for obtaining feedback in systems incorporating explicit relevance feedback, whether by requiring the user to assign specific numeric ratings to an object or by requiring the user to indicate if the object is good or bad.

Behavior category (examine, retain, reference, annotate and create) refers to the underlying intent of the observed behavior. Examine are those behaviors that the user engages in to further investigate an information object. Examples of examine behaviors are view (i.e. reading time), listen and select. Retain are those behaviors that suggest some degree of intention to make future use of an object. Examples of retain behaviors are bookmark, save, and print. Reference are those behaviors that function to establish some form of a link between two objects and include the behaviors of reply, link and cite. Finally, annotate are those behaviors that the user engages in to intentionally add personal value to an information object. Example behaviors from this category include mark-up, rate and organize. (Create is described below.)

Minimum scope (segment, object, and class) refers to the smallest possible scope of the item being acted upon. A segment is a portion of a larger information object. For instance, a paragraph or page is a segment of a document. Objects are self-contained items such as documents, videos or images. A class is a group of objects. A collection of indexed documents, such as an online database like PsycInfo, can be considered as a class of objects. These categories are not orthogonal and should be viewed instead as nested information units (i.e. segments comprise objects, which in turn, comprise classes).

Kelly and Teevan (2003) further developed the framework proposed by Oard and Kim (2001) by adding a fifth behavior category, *Create*, which describes those behaviors the user engages in when creating original information. Examples of create behaviors include type, edit and author. Type and edit behaviors are reserved for situations where the user adds original text to a preexisting object. The behavior of *author* is reserved for situations where the user creates an original object from scratch, such as a paper or an email message. In addition to these behaviors, Kelly and Teevan (2003) added several other behaviors to the categories of *examine* and *retain*. Additions made to the *examine* category include scroll, find, query and browse. Additions made to the *retain* category include email, where email describes the behavior where one finds a useful document and emails it to oneself.

Categorizing an observable behavior into the appropriate cell in Figure 1 can be difficult, because both the intent of the behavior and its scope can be ambiguous. For example, the behavior of saving a newly created document is obviously a retention behavior, but the user also has to create a label (i.e. title) for the document. Furthermore, the user has to save to document to a particular location, and thus performs an annotation behavior (organize). This behavior involves the user organizing objects, either those personally created or those obtained from other sources, into folders (for instance). While the act of saving these objects can be viewed as a type of retention or creation, the structure that the user imposes on the objects by grouping them together or separately, can be viewed as a type of annotation.

It can also be problematic assigning a minimum scope to a behavior because the scope is often ambiguous. For example, a behavior such as bookmark acts on a Web page, which is traditionally considered an object. However, when a Web page is considered in the context of its containing Web site, it can be understood as a segment instead. Note, too, that observable behaviors are classified according to the minimum scope for which the behavior could be observed. For example, the minimum scope for which one might observe the behavior *type* is *segment* although it is also common for typing to occur during the creation of an object. Similarly, *view* is identified in the *examine segment* category. However, most research has investigated viewing as it relates to objects, and thus that research would be categorized in the *examine object* category.

A final distinction should be made regarding the classification of the behaviors of *find* and *query* into the examine category. While find and query behaviors involve the creation of text, they are primarily used to locate information for examination, and are thus classified in the *examine* category. For example, a user might use *find* to locate a term or passage to examine in a document. Similarly, a user might perform *querying* to locate a document for examination. While querying traditionally applies to documents, the behavior is classified with a minimum scope of segment because some systems return best passages rather than documents.

One interesting thing that this classification highlights is the differences in precision that accompanies minimum scope and in the confidence one can have with each class of behaviors. For instance, obtaining implicit feedback about a segment will presumably provide more precise information about the user's interests than obtaining implicit feedback about a class of objects. Additionally, confidence in the evidence that is available for inference varies according to observed behavior. In part, confidence is related to the number of opportunities that one has to observe behaviors and the user's deliberateness in engaging in such behaviors. Examination behaviors are weaker evidence than annotate or create behaviors since they occur for every object with which the user interacts. Conversely, annotate or create behaviors provide stronger evidence since they occur less frequently and more deliberately.

It is further worth noting the impact of interface design and human-system interaction style in determining what types of implicit feedback are available for observation and use. For instance, the system CiteSeer (Bollacker, Lawrence, and Giles, 2000), an automatic generator of scientific literature databases, first displays the document citation to the user, who is then given several options, including view the full text, rate the document, view citations made to the document and view the bibliography of the document. This type of interaction style provides more opportunities to collect implicit and explicit feedback than one which only allows the user to query and examine search results. As we learn more about the value of implicit feedback techniques, it will be interesting to see if this translates into new and different interfaces features and interaction styles to support implicit feedback.

Finally, there is one major class of behaviors that has been intentionally omitted from Figure 1, those behaviors which can only be observed with specialized hardware, such as eye movement, gesture and mouse pressure. There is a large and growing body of research in the human-computer interaction community using these types of behaviors to infer things such as attention (Fendlay, Walker, and Kentridge, 1995), and more recently, cognitive load (Ikehara, Chin, and Crosby, 2003) and emotion (Picard and Klein, 2002). While it is possible that information obtained from these types of specialized hardware can provide useful implicit feedback for IR related tasks, this discussion does not consider these classes of behaviors.

3. IMPLICIT FEEDBACK RESEARCH

Implicit feedback techniques have been used to retrieve, filter and recommend a variety of items: hyperlinks, Web documents, academic and professional journal articles, email messages, Internet news articles, movies, books, television programs, jobs, and stocks. Two of the most widely investigated behaviors for implicit feedback are object (i.e., document) selection and viewing time, both of which are included in the examine object category of the classification in Figure 1. Because these behaviors are seemingly easy to monitor and gather and are available for every object with which the user interacts, many researchers have elected to study these behaviors. There is less research across the minimum scope categories of segment and class because for most systems, the unit with which the user most often interacts is the object. Furthermore, there is less research investigating the behaviors of retain, reference, annotate and create since it is often necessary to collect this information from the client, rather than the server, which usually requires specialized software and permission from users.

Implicit feedback research can be described according to its research approach and to the degree of personalization that it offers users. From a research approach, implicit feedback research can be understood as theoretical, basic or applied. Research that is theoretical proposes models of how systems using implicit feedback ought to work. This type of research includes those pieces that describe the design of systems using implicit feedback. Basic research identifies various potential sources of implicit feedback, evaluates how well the sources do in predicting a user's interests and develops formalizations or rules describing this relationship. Applied research builds, implements and/or evaluates systems or algorithms which use implicit feedback techniques to construct user profiles and retrieve, filter or recommend information objects. The research on implicit feedback is quite evenly distributed over these types of approaches. Implicit feedback research can further be described according to the degree of personalization it offers users. Degree of personalization can be considered as a function of the techniques used to explore the relationship between document preference and behavior, and the techniques used to filter, recommend or retrieve objects. Techniques used to explore the relationship between document preference and behaviors can be based on an examination of a single user's behaviors, based on the examination of a sample of users' behaviors or on a hybrid approach using both methods.

For example, the amount of time an individual spends viewing a document can be compared to that individual's explicit preferences for documents to understand if and how viewing time can be used as an implicit measure for preference for that particular user, or viewing times can be averaged across a sample of users and document preferences to arrive at a general, all-purpose rule for using viewing time as an implicit measure of preference. Techniques used to recommend, retrieve or filter information objects can be personal, as when the system makes recommendations to a single user, or these techniques can be collective, as when the system provides feedback on an aggregate level by, for example, clustering the documents according to access patterns or highlighting popular articles.

Finally, these techniques can be a hybrid of the two, making personal recommendations when enough information about the user is available and making collective recommendations when little or no information is available. Most research on behaviors as implicit feedback has explored the former two approaches; little work has explored hybrid approaches.

Almost all of the research on implicit feedback has considered how behaviors can be used as positive evidence, rather than negative evidence. However, one can imagine behaviors which indicate that a user does not find something relevant or which suggest that something is unimportant to the user, such as delete. It is likely that little research has been conducted on negative implicit feedback because there are fewer of these types of behaviors, and, in general, less is understood about how to effectively use negative feedback, whether for implicit or explicit relevance feedback.

A final comment about how evaluations of behavior as implicit feedback have been conducted is worth mentioning. In many studies of implicit feedback (Cooper and Chen, 2002), retention behaviors (e.g., printing, saving, and emailing) are assumed to indicate relevance and are used as the benchmark with which to evaluate implicit feedback. This is in contrast to studies that consider these same behaviors as potential measures of implicit feedback and evaluate such behaviors against user judgments of document relevance. Indeed, studies of behavior as implicit feedback have used a variety of techniques to establish a benchmark with which to evaluate behaviors; techniques have ranged from asking users to grade documents (using A, B, C, F), to providing users with Likert-type scales for assessing relevance. This diversity is one possible reason why research in this area has had mixed results. Unfortunately, in many cases, how users evaluate the relevance of documents has only been superficially considered (if considered at all) and the extensive body of research on relevance judgments and measurement has been largely ignored.

A selection of research on implicit feedback is presented below to illustrate how such studies have been conducted, how feedback has typically been measured and used, and the key issues and problems associated with implicit feedback. This research is presented according to its behavioral category as identified in Figure 1; however, many papers discuss techniques using behaviors from more than one category. It is difficult to compile an exhaustive list of the literature on implicit feedback as it can be found in numerous publication types, across a variety of specialty areas including: user modeling, data mining, machine learning, intelligent user interfaces, human-computer interaction, e-commerce, IR and filtering, recommendation systems, and general web studies. Research on implicit feedback is also commonly referred to as web-usage mining, especially in the data mining community. The research that is presented in this chapter is limited to the use of implicit feedback techniques for IR related tasks.

3.1. Examine

Much of the research on using behavior as implicit feedback for IR related tasks has investigated behaviors from the examine behavior category. This category includes the behaviors of view, listen, scroll, find, query, select and browse. It should be noted that 'view' and 'read' are often used interchangeably and treated as equivalents to the length of time that an object is displayed. Of course, the amount of time that an object is displayed does not necessarily correspond to the amount of time that the object is viewed.

Moreover, treating the length of time a document is displayed as equivalent to the length of time the document is read is a huge conceptual leap, given the complexity of a behavior like reading. For instance, a vast majority of reported 'reading' times are quite low (e.g., less than one minute), so one could argue that users are not engaged in reading per se, but rather some combination of quick skimming and decision-making. Furthermore, researchers have demonstrated no direct correlation between document length and display time (Miller, Riedl, and Konstan, 2003; Morita and Shinoda, 1994). Indeed, much of the research on behavior as implicit relevance feedback makes some assumptions about what observable behaviors, which are usually 'observed' through a system log, represent. This is an outstanding problem, one which some are beginning to address (Kelly and Belkin, 2004). In presenting the work below, behaviors will be referred to as they are by the author(s) of the work discussed.

Claypool et al. (2001) provided a categorization of different interest indicator categories, both explicit and implicit, and addressed the fundamental question of which observable behaviors could be used as implicit measures of interest. Several behaviors were examined: mouse clicks, scrolling, and time on page. Mouse clicks and scrolling were measured both as a frequency number (i.e. number of mouse clicks) and as total time spent. Scrolling was further measured both at the keyboard and with the mouse. Users were asked to explicitly rate each page that they viewed just before the page closed and these ratings were used to evaluate the implicit measures. The authors found that time spent on a page, the amount of scrolling on a page (all scrolling measures combined) and the combination of time and scrolling had a strong positive correlation with the explicit ratings.

Morita and Shinoda (1994) explored how behaviors exhibited by users while reading articles from newsgroups could be used as implicit feedback for profile acquisition and filtering. The authors measured the behaviors of reading time, saving, following-up and copying for each of the 8,000 articles read by their eight users. They further examined the relationship of three variables on reading time: the length of the document, the readability of the document and the number of news items waiting to be read in the user's news queue. Very low correlations were found between the length of the article and reading time, the readability of an article and reading time and the size of the user's news queue and reading time. Although no statistics were presented, the reading time for articles rated as interesting was longer than for articles rated as uninteresting. Saving, following-up and copying of an article were not found to be related to interests.

The finding that users spend more time reading documents that they find relevant has been replicated a number of times in a variety of settings (Cooper and Chen, 2002; Miller, Riedl, and Konstan, 2003; Seo and Zhang, 2000), although there are some limitations to its usefulness as an implicit measure. For instance, Miller, Riedl, and Konstan (2003) found that in Usenet news groups, users spent longer reading news items they found relevant, than those they did not find relevant, although the correlation was not high (0.36). The correlation between explicit ratings of relevance and time spent reading was further examined as it related to each user. Miller, Riedl, and Konstan (2003) found a range of correlations, both positive and negative, from −.50 to .80, suggesting that reading time does not predict relevance equally as well, or in the same direction, for all individuals. Kelly and Belkin (2001) found similar results.

Furthermore, it is unclear how such information can be most effectively used, especially in real-time retrieval situations. For instance, in a controlled laboratory study, White, Ruthven, and Jose (2002a) examined reading time as a technique for automatically re-ranking sentence-based summaries for retrieved documents. Users completed simulated tasks using three types of systems, one of which automatically re-ranked the top sentences in the summaries based on the user's reading time of each summary. Performance results regarding this system were inconclusive. This group has continued to work in this area and has had some success with using implicit feedback in real-time retrieval situations (White, Jose, and Ruthven, 2003; White and Jose, 2004).

Finally, listening time has been explored as a source of user preference for news stories distributed via an IP-enabled radio (Billsus and Pazzani, 1999). Similar to document filtering, the assumption is that users will listen longer to those stories they consider interesting as opposed to those that they do not consider interesting. Billsus and Pazzani (1999) evaluated the validity of this proposition by retroactively using listening time as a measure of user preference and comparing it to explicit ratings of stories. The results indicated that when used with explicit ratings, listening time can provide additional information about user preference for stories. It should be noted that as a measure of user preference, listening time is more sensitive than reading time since it is possible to measure how much of the document the user has heard.

3.1.1. Scrolling

Scrolling is a behavior that has received less attention as a source of implicit feedback because of the difficulty in obtaining an accurate measure of its occurrence. Scrolling must be captured on the client-side and because scrolling can occur via multiple methods (up and down arrow keys, page up and down keys, mouse clicks on the scroll bar, mouse wheeling and dragging the scroll bar) it is often difficult to design a logger which captures all of these methods. However, results of analysis using scrolling as a predictor of relevance demonstrate some potential.

Claypool et al. (2001) found that all scrolling methods combined were good predictors of relevance, but were not so when examined alone. Kamba, Sakagami, and Koseki (1997) implemented a system which presented users with news articles and used scrolling, displaying in a separate window and magnifying an article as implicit sources of feedback. The absence of any of these behaviors was further used as negative feedback. Two version of this system, one which increased and decreased scores by 10 points and another by 30 points, were compared to a system where users provided explicit feedback.

The results of the study found that the difference in anticipated relevance scores and actual relevance scores decreased quickly in the version of the system with explicit feedback, less quickly in the version of the implicit feedback system which changed scores by 10 points and not at all in the system which changed scores by 30 points. Kamba, Sakagami, and Koseki (1997) observed that their use of negative feedback was neither sensitive nor selective enough, and negatively impacted the performance of the system, especially since their users often did not perform any activity on relevant documents and performed many activities on non-relevant documents.

3.1.2. Selection

Selection, or click-through, is a behavior that has received a lot of attention, especially in the machine learning and data mining communities. This behavior is very easy to obtain from proxy servers and exists in large quantities. However, it should be noted that selection is a weaker form of implicit feedback than reading time or scrolling, since it does not distinguish between useful and non-useful Web pages requested by a user. Instead, it assumes that all pages requested by the user are equally useful to the user. Anyone who has had experience using the Web can see that this assumption is problematic. Nonetheless, researchers have had some success recommending items to users based on selection or click-through data. Click-through data has also been used to re-rank search results (Joachims, 2002), identify preferred categories of interests for users (Chen, LaPaugh, and Singh, 2002), and as an advertising revenue model for search engines such as Google.

Joachims, Freitag, and Mitchell (1997) introduced WebWatcher that accompanied users browsing a specific website and recommended pages to that user based on the user's previous selection behaviors and the selection behaviors of others who had previously browsed the same information. Personal WebWatcher (Mladenic, 1999) expanded WebWatcher by building a user model by performing a content analysis on Web pages that a single user requested. The content analysis resulted in a list of extracted key terms that were used to describe the user's preferences; this list was then used to identify and recommend links that were believed to be of interest to the user. Letizia (Lieberman, 1995) performed a breadth search on the contents of a user's selected Web pages and suggested potentially relevant hyperlinks based on this analysis.

3.1.3. Querying

The use of querying behavior as a source of implicit feedback has recently gained some attention as a source for information-need context and as a source of terms for query expansion. For example, Huang, Chien, and Oyang (2003) developed a log-based approach to identifying, extracting and suggesting terms for query expansion based on query sessions with similar requests. Query expansion terms were extracted from queries that were determined to be similar to one another; thus, users functioned as sources for term suggestion.

Shen and Zhai (2003) proposed a technique for using a user's query history to expand their current query. Results from this study demonstrated that retrieval performance could be consistently improved using this approach. Liu, Yu, and Meng (2002) created personalized Web searches by mapping user queries to categories. All of these studies suggest the potential of query history as a source of implicit feedback for query expansion, term disambiguation and profile acquisition.

3.2. Retain

There is less research on the effectiveness of retention behaviors as sources of implicit feedback. Retention behaviors include saving, printing and book marking. Because these behaviors are performed locally by the user, they must either be gathered from the client machine or extensive work must be performed server-side to collect the data. Thus, these behaviors have been primarily studied in controlled laboratory experiments.

The focus of studies of retention behaviors has typically been on printing. For instance, Kim, Oard, and Romanik (2000) examined the relationship between reading time and printing and a user's interest in a document. This experiment examined this relationship in the context of academic and professional journal articles, which users accessed via Dialog. In two studies, Kim, Oard and Rominik (2000) did not find a linear relationship between reading time and relevance. Instead, they found that mean reading time for non-relevant documents was lower than that for moderately relevant or relevant. However, the mean reading time for relevant documents was lower than that for moderately relevant, which suggests that users spend longer reading documents when the relevance of the document is not obvious.

This may be because of a lack of understanding of their information need or the contents of the document. For printing behavior, Kim, Oard, and Romanik (2000) found that 16 of the 153 documents examined by users were printed and that all printed documents were identified as moderately or highly relevant by users. The mean reading time of printed documents was only 2.28 seconds more than that of non-relevant documents; thus, it appears that printing behavior can be used as an independent source of implicit feedback.

Book marking has also been used as a source of implicit feedback. Rucker and Polanco (1997) used bookmarks and their organization to identify communities of interests and to recommend Web pages by computing the degree of overlap between different users' bookmarks. The organization of bookmarks into folders was used as evidence of the semantic coherence of a category of interest and as a source of relevant groupings between individuals. In a study of over 1000 users, it was found that this technique worked best on folders that had between 15 and 20 bookmarks. Seo and Zhang (2000) evaluated reading time, scrolling, link selection and book marking as implicit measures of interest and found that book marking was the most strongly related to interest; reading time and link selection were also found to be related, but less strongly, and scrolling was not found to be related to relevance.

3.3. Reference

Perhaps the most impressive large-scale use of implicit feedback comes in the form of Web link analysis, which uses the reference behavior, link, as its source of data. An example of this is Kleinberg's (1999) work with hubs and authorities. *Authorities* are authoritative information sources on a topic, and *hubs* are collections of authorities. Kleinberg (1999) suggested that good hubs could be recognized because they point to many good authorities, and similarly, good authorities could be recognized because they are pointed to by many hubs. Thus the links that people make in the course of Web page authoring are interpreted as implicit feedback. Link analysis, in the form of PageRank (Page et al., 1998) is used to great success in practice by Google.

Infoscope (Fischer and Stevens, 1991) provided users with an agent that helped them create a personalized Boolean filter for email messages. While users interacted with their email messages, the agent monitored behaviors such as saving, deleting and replying to identify patterns of behavior. The agent used this information to model the user's interactions with email and suggested possible filtering preferences to the user. Goldberg et al. (1992) proposed a content- and collaborative-based email filtering system that used replying as one of its major sources of information about the user. Finally, Terveen et al. (1997) created a system to recommend URLs in Usenet news groups. Recommendations were based on previous postings made by individual subscribers to the news groups; URLs that were referred to or mentioned in these posting were identified as potential candidates for future recommendations.

3.4. Annotate

The annotation behavior of users as implicit feedback has not received much attention, since wide-spread use of systems supporting annotation is uncommon. While some systems create artificial opportunities for users to enter annotations via keyboard, these types of systems are not considered here since they require the user to explicitly make annotations. Instead, only systems which use the annotations made by users as indirect feedback are discussed.

Golovchinsky, Price, and Schilit (1999) constructed full text queries based on users' annotated passages of documents and compared these to queries constructed using standard relevance feedback techniques. This work was motivated by the observation that the words and passages that users mark can provide the system with a more refined, user-specific unit with which to perform relevance feedback, and that these passages can help in establishing search context. Results from an experiment with ten users annotating and evaluating documents for six topics found that queries derived from users' annotations resulted in retrieval performance that was better than standard relevance feedback techniques.

Price, Golovchinsky, and Schilit (1998) used the text located near free-form annotations made by users to retrieve new documents, and displayed links to these new documents to users in the margin of the document that they were currently reading and annotating. Further work on the use of annotation behaviors as sources of implicit feedback is described in Shipman et al. (2003).

3.5. Create

The final behavioral category from Figure 1 is create. Create considers the behaviors of type, edit and author. Similarly to the research on annotate, there is little that has been done on using create behaviors as sources of implicit feedback. However, unlike annotate behaviors, create behaviors are more readily available since word processing is a common application. Additionally, create behaviors provide a potentially richer source of implicit evidence, since the text that a user creates represents, to a certain extent, that user's understanding of a particular topic or problem. However, since these activities are performed on a local machine it can be difficult to collect this data. Furthermore, it is even more difficult to effectively use and make sense of this data since the unit in which it comes can be quite large.

Budzik, Hammond, and Birnbaum (2001) and Budzik and Hammond (1999) proposed Watson, a system that automatically retrieved documents and recommended URLs to users based on what users typed. This work was motivated in part by the observation that users typically pose short queries that are highly ambiguous and often lack context. Budzik and Hammond (1999) suggested that evidence of context can be found in numerous other applications with which the user interacts. To provide some support for their hypothesis, the authors asked ten researchers to submit an electronic version of a paper that they wrote and then asked these users to evaluate the documents that their experimental system had retrieved based on these texts. The results were encouraging with at least eight of the ten users indicating that at least one of the retrieved results would have been useful. While Budzik and Hammond (1999) also provided results from several other informal evaluations, a full-scale, formal evaluation has yet to be performed.

Rhodes and Starner (1996) proposed a system, Remembrance Agent, which watched what the user typed and read, and then found and displayed old email and notes from the user's personal files that it believed were related to the user's current activities. Rhodes (2000) followed-up Remembrance Agent with Margin Notes, a system that examined the document that a user was in the process of creating, and recommended other documents, including online documents, based on their similarity to the current document. In more recent work, Dumais et al. (2004) developed a system that issues 'implicit queries' where queries are automatically generated based on users' current computing activities and results are presented in the context of users' ongoing work.

4. CHALLENGES FOR FUTURE WORK

Implicit feedback techniques offer substantial promise with regard to automatically capturing and modeling user preferences and performing personalized retrieval, filtering and recommendation. However, the review of research has demonstrated several limitations of these techniques. There are numerous problems related to how these behaviors are collected, measured, interpreted and applied. These issues include those that are related to information-seeking context, relevance, degree of personalization and measurement.

It is clear that numerous problems arise when trying to infer information from observable behaviors because what can be observed does not necessarily reflect users' underlying intentions. For instance, the amount of time that an object is displayed does not necessarily correspond to the amount of time that the object is examined, yet display time traditionally is treated as an equivalent to reading time. Further, the amount of time that an object is actively examined does not necessarily correspond to the user's interest in that object; instead this behavior might be a result of some other factor such as the user's previous experience with, and knowledge of, the topic.

Finally, not all behaviors are equally as useful as indicators of relevance and some may only be useful in combination with others. For instance, the selection of an object is weaker evidence of a user's interests than printing or saving since the former behavior occurs for every object with which the user interacts, and the latter does not. Additionally, the use of behavior as implicit relevance feedback could potentially generate conflicting evidence, as in the case of a document that is displayed for a short amount of time and printed.

While it appears that implicit feedback can be useful, in most cases it is not inherently so and should be understood within the context in which it occurs. Information-seeking research has demonstrated that the relationship between behaviors and relevance can be confounded by numerous contextual factors such as task and topic, as well as problem solving stage and time. For instance, task has been used to explain differences in relevance assessments of information objects, and differences in other searching behaviors, such as selection of search tactics and terms (Vakkari, 2003). Kuhlthau (1993) demonstrated that users progress through a series of information-seeking stages, adopting different strategies and exhibiting different behaviors over time. Vakkari (2000) found that the degree of students' knowledge of their paper topic affected their ability to express search terms and formulate search tactics.

Although early research on the use of behavior as implicit feedback did not draw heavily on information-seeking research, more current studies have demonstrated the importance of this research. For instance, Kelly and Cool (2002) examined the relationship between display time, relevance and topic familiarity and found that display time varied according to relevance and topic familiarity. Results demonstrated that higher display times were not only associated with more highly relevant documents, but that they were also associated with lower levels of topic familiarity. These results suggest the occurrence of some interaction between familiarity, usefulness, and display time. In a naturalistic study, Kelly and Belkin (2004) found that display time varied considerably according to specific task and according to specific user.

Kellar et al. (2004) examined the relationship between task, time spent reading and relevance in a controlled laboratory setting and found that the effectiveness of reading time as a measure of relevance was related to task. Specifically, Kellar et al. (2004) found that reading time was a more effective measure of relevance for complex tasks, where users were asked to make a comparison within a document or complete a small calculation using data from within the document. Thus, to most effectively use behavior as implicit relevance feedback more research needs to be conducted on understanding how behaviors change with respect to contextual factors.

The research on implicit feedback has also done an inadequate job of incorporating the large body of research on relevance. Although the general idea behind the use of implicit feedback is to personalize IR interactions to the individual user, the dynamic and multidimensional character of relevance has not been acknowledged. The research on relevance suggests that it would be difficult to consistently infer the usefulness of an information object based on behavior without considering some user- and/or context-specific attributes of the information-seeking situation.

For instance, Spink (1996) found that at the initial stage of problem solving, people tended to judge more documents as partially relevant than as fully relevant. Quiroga and Mostafa (2002) identified several characteristics of the user's background and information needs, and of the document collection that influenced relevance feedback judgments in a study aimed to clarify the role of context in relevance feedback judgments. Previous research using behavior to infer relevance has failed to consider the findings from this research and has instead considered relevance judgments as somewhat stable over time and across situations.

Another challenge facing implicit feedback research is the notion of degree of personalization offered by the system. In particular, individual differences can greatly impact the effectiveness of using behavior as implicit relevance feedback. People behave differently and have varying approaches to information-seeking; thus, it is difficult to generate, and dangerous to apply, all-purpose rules for describing how behavior can be used as implicit relevance feedback. For instance, Kelly and Belkin (2001) examined the relationship between reading time and relevance for thirty-six users engaged in six different IR tasks. While overall Kelly and Belkin (2001) were able to find a statistically significant difference in reading time for relevant and non-relevant documents, this relationship held for only one-third of the study participants and for several participants, this relationship was in the direction opposite of what was expected.

These findings are similar to those of Miller, Riedl, and Konstan (2003). This research suggests that the degree of personalization is an important issue to consider when using behavior as implicit feedback. Degree of personalization is likely to change with respect to the purposes of the service and with the anticipated diversity or homogeneity of users. It is likely that the best approaches to using behavior as implicit feedback will incorporate hybrid techniques incorporating both general and personal behavioral models. For instance, general behavioral models might function to describe how behavior can potentially be used as implicit feedback and provide starting points upon which to begin the construction of personal behavioral models (Kelly and Belkin, 2002).

A final challenge for future research is to better understand the relationship between various metrics used as implicit feedback and the behaviors that they are meant to represent, as well as the techniques used to collect and compute these metrics. Studies of implicit feedback have spent little time exploring the validity and reliability of the metrics that they employee, and have instead assumed that the metrics represent the actual behavior of interest (e.g., the length of time a document is displayed in the browser window is equivalent to how long the user reads the document); and that the measurement technique is valid and reliable (e.g., display time collected at the proxy represents the actual time the document is displayed in the user's browser window, and this measurement is consistent regardless of document type or network speed). Moreover, while these studies often describe where metrics were collected (proxy or client), they do not describe how metrics were computed. For example, display time is not directly available at the proxy. Instead, it is usually computed from other available information, such as the difference between successive URL requests. Understanding how such metrics are computed is important in evaluating the quality of the metric.

Furthermore, Kelly and Belkin (2004) demonstrated that the integrity of behavior-based metrics used in studies of implicit feedback is an important research issue when they found that proxy-generated display times were neither valid nor reliable substitutes for client-generated display times. Although studies based exclusively on proxy-generated data are attractive since a great deal of data can be collected in a relatively short period of time with little cost, just how behavior is measured is likely to strongly influence its utility; simple, easily collected proxy-side data may well give misleading results.

5. CONCLUSIONS

The chapter presented a framework for understanding and studying how behavior can be used as implicit relevance feedback. This included a classification and discussion of behaviors that have been used as implicit relevance feedback, a general discussion and characterization of implicit feedback research, and a presentation of example studies to illustrate how such studies have been conducted and how feedback has typically been measured and used. Finally, this chapter presented a discussion of key issues and problems associated with implicit feedback research and identified challenges for future research. The use of behavior as implicit relevance feedback is an exciting and promising approach to personalizing IR interactions. Although more effort needs to be made to fully understand how behaviors can be used as implicit relevance feedback, current research efforts offer a promising start.

REFERENCES

Anick, P. (2003). Using Terminological Feedback for Web Search Refinement: A Log Based Study. In *Proceedings of the 26th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR '03)* (pp. 88–95), Canada.

Beaulieu, M. (1997). Experiments on Interfaces to Support Query Expansion. *Journal of Documentation,* 53(1), 8–19.

Belkin, N. J., Cool, C., Kelly, D., Lin, S. J., Park, S. Y., Perez-Carballo, J., and Sikora, C. (2001). Iterative Exploration, Design and Evaluation of Support for Query Reformulation in Interactive Information Retrieval. *Information Processing and Management,* 37(3), 404–434.

Billsus, D. and Pazzani, M. J. (1999). A Personal News Agent That Talks, Learns and Explains. In *Proceedings of the 3rd International Conference on Autonomous Agents (AGENTS '99)*, (pp. 268–275), USA.

Bollacker, K. D., Lawrence, S., and Giles, C. L. (2000). Discovering Relevant Scientific Literature on the Web. *IEEE Intelligent Systems,* 15(2), 42–47.

Borlund, P. (2003). The Concept of Relevance in IR. *Journal of the American Society for Information Science,* 54(10), 913–925.

Budzik, J. and Hammond, K. (1999). Watson: Anticipating and Contextualizing Information Needs. In *Proceedings of the 62nd Meeting of the American Society for Information Science (ASIS '99)*, (pp. 727–740), USA.

Budzik, J., Hammond, K., and Birnbaum, L. (2001). Information Access in Context. *Knowledge Based Systems,* 14(1–2), 37–53.

Chen, M. LaPaugh, A. S. and Singh, J. P. (2002). Predicting Category Access for a User in Structured Information Space. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*, (pp. 65–72), Finland.

Claypool, M. Le, P. Waseda, M. and Brown, D. (2001). Implicit Interest Indicators. In *Proceedings of the 6th International Conference on Intelligent User Interfaces (IUI '01)*, (pp. 33–40), USA.

Cooper, M. D. and Chen, H.-M. (2001). Predicting the Relevance of a Library Catalog Search. *Journal of the American Society for Information Science,* 52(10), 813– 827.

Dumais, S., Cutrell, E., Sarin, R., and Horvitz, E. (2004). Implicit Queries (IQ) for Contextualized Search. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'04)*, Sheffield, UK, 594.

Fendlay, J. M., Walker, R., and Kengridge, R. W. (1995) (Eds.). *Eye Movement Research: Mechanisms, Processes and Applications*. New York: Elsevier Science Publishing.

Fischer, G. and Stevens, C. (1991). Information Access in Complex, Poorly Structured Information Spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*, (pp. 63–70), USA.

Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM,* 35(12), 61–70.

Golovchinsky, G., Price, M. N., and Schilit, B. N. (1999). From Reading to Retrieval: Freeform Ink Annotations as Queries. In *Proceedings of the 24th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, (pp. 19–25), USA.

Huang, C.-K., Chien, L.-F., and Oyang, Y.-J. (2003). Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs. *Journal of the American Society for Information Science and Technology,* 54(7), 638–649.

Ikehara, C. S., Chin, D. N., and Crosby, M. E. (2003). A Model for Integrating an Adaptive Information Filter Utilizing Biosensor Data to Assess Cognitive Load. In *Proceedings of the Ninth International Conference on User Modeling (UM '03)*, (pp. 208–212), USA.

Joachims, T. (2002). Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD '02)*, (pp. 133–142), Canada.

Joachims, T. Freitag, D., and Mitchell, T. (1997). WebWatcher: A Tour Guide for the World Wide Web. In *Proceedings of the International Joint Conference on Artificial Intelligence (ICML 97)*, (pp. 770–777), USA.

Kamba, T., Sakagami, H., and Koseki, Y. (1997). ANATAGONOMY: A Personalized Newspaper on the World Wide Web. *International Journal of Human-Computer Studies,* 46, 789–803.

Kellar, M., Watters, C., Duffy, J., and Shepherd, M. (2004). Effect of Task on Time Spent Reading as an Implicit Measure of Interest. In *Proceedings of the 67th American Society for Information Science and Technology (ASIS and T) Annual Meeting,* 41, pp. 168–175.

Kelly, D. and Belkin, N. J. (2001). Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback During Interactive Information Retrieval. In *Proceedings of the 24th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR '01)*, (pp. 408–409), USA.

Kelly, D. and Belkin, N. J. (2002). A User Modeling System for Personalized Interaction and Tailored Retrieval in Interactive IR. In *Proceedings of the Annual Conference of the American Society for Information Science and Technology (ASIST '02)*, (pp. 316–325), USA.

Kelly, D. and Belkin, N. J. (2004). Display Time as Implicit Feedback: Understanding Task Effects. In *Proceedings of the 27th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '04)*, (pp. 377–384), Sheffield, UK.

Kelly, D. and Cool, C. (2002). The Effects of Topic Familiarity on Information Search Behavior. In *Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL '02)*, (pp. 74–75), USA.

Kelly, D. and Teevan, J. (2003). Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum* 37(2), 18–28.

Kim. J. Oard, D. W. and Romanik, K. (2000). *Using Implicit Feedback for User Modeling in Internet and Intranet Searching.* University of Maryland CLIS Technical Report 00-01.

Kleinberg, J. M. (1999). Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM,* 46(5), 604–632.

Koenemann, J. and Belkin, N. J. (1996). A Case for Interaction: A Study of Interactive Information Retrieval Behavior and Effectiveness. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '96)*, (pp. 205– 212), Canada.

Kuhlthau, C. C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex.

Lieberman, H. (1995). Letizia: An Agent That assists Web Browsing. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 924–929), USA.

Liu, F., Yu, C., and Meng, W. (2002). Personalized Web Search By Mapping User Queries to Categories. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM '02)*, (pp. 558–565), USA.

Miller, B. N., Riedl, J. T., and Konstan, J. A. (2003). GroupLens for Usenet: Experiences in Applying Collaborative Filtering to a Social Information System. In: C. Lueg and D. Fisher (Eds.), *From Usenet to CoWebs: Interacting With Social Information Spaces* (pp. 206–231) London: Springer Press.-

Mladenic, D. (1999). Machine Learning Used By Personal WebWatcher. *Proceedings of the ACAI-99 Workshop on Machine Learning and Intelligent Agents*. Crete.

Morita, M. and Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '94)*, (pp. 272–281), Ireland.

Nichols, D. M. (1997). Implicit Ratings and Filtering. In *Proceedings of the 5th DELOS Workshop on Filtering and Collaborative Filtering* (pp. 31–36), Hungary.

Oard, D. W. and Kim, J. (2001). Modeling Information Content Using Observable Behavior. In *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology (ASIST '01)*, (pp. 38–45), USA.

Page, L. Brin, S. Motwani, R. and Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web.* Stanford University Technical Report.

Picard, R. W. and Klein, J. (2002). Computers That Recognize and Respond to User Emotion: Theoretical and Practical Implications. *Interacting with Computers,* 14(2), 141–169.

Price, M. N., Golovchinsky, G., and Schilit, B. N. (1998). Linking By Inking: Trailblazing in a Paper-Like Hypertext. In *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia '98 (Hypertext '98)*, (pp. 30–39), USA.

Quiroga, L. M. and Mostafa, J. (2002). An Experiment in Building Profiles in Information Filtering: The Role of Context of User Relevance Feedback. *Information Processing and Management,* 38, 671–694.

Rhodes, B. J. and Starner, T. (1996). Remembrance Agent: A Continuously Running Automated Information Retrieval System. In *Proceedings of the 1st International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology (PAAM '96)*, (pp. 487–495).

Rhodes, B. J. (2000). Margin Notes: Building a Contextually Aware Associative Memory. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI '00)*, (pp. 219–224), USA.

Rucker, J. and Polanco, J. (1997). Citeseer: Personalized Navigation for the Web. *Communications of the ACM,* 40(3), 73–75.

Salton, G. (Ed.) (1971). *The SMART Retrieval System: Experiments in Automatic Document Processing*. NJ: Prentice-Hall.

Seo, Y. W. and Zhang, B. T. (2000). A Reinforcement Learning Agent for Personalized Information Filtering. In *Proceedings of the 5th International Conference on Intelligent User Interfaces (IUI '00)*, (pp. 248–251), USA.

Shen, X. and Zhai, C.-X. (2003). Exploiting Query History for Document Ranking in Interactive Information Retrieval. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03)*, (pp. 377–378), Canada.

Shipman, F. M., Price, M. N., Marshall, C. C., Golovchinsky, G., and Schilit, B. N. (2003). Identifying Useful Passages in Documents Based on Annotation Patterns. In *Proceedings of the 2003 European Conference on Digital Libraries (ECDL '03)*, (pp. 101–112).

Spink, A. (1996). Multiple Search Session Model of End-User Behavior: An Exploratory Study. *Journal of the American Society for Information Science,* 47(8), 603–609.

Spink, A. and Losee, R. M. (1996). Feedback in Information Retrieval. *Annual Review of Information Science and Technology,* 31, 33–78.

Terveen, L. Hill, W. Amento, B. McDonald, D. and Crete, J. (1997). PHOAKS: A System for Sharing Recommendations. *Communications of the ACM,* 40(3), 59–62.

Vakkari, P. (2000). Cognition and Changes of Search Terms and Tactics During Task Performance: A Longitudinal Case Study. In *Proceedings of the RIAO 2000 Conference*, 894–907.

Vakkari, P. (2003). Task-Based Information Searching. *Annual Review of Information Science and Technology,* 37, 413–464.

White, R. W. Ruthven, I. and Jose, J. M. (2002a). Finding Relevant Documents Using Top Ranking Sentences: An Evaluation of Two Alternative Schemes. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02)*, (pp. 57–64), Finland.

White, R. W. Jose, J. M. and Ruthven, I. (2003). An Approach for Implicitly Detecting Information Needs. In *Proceedings of the 12th Annual ACM Conference on Information and Knowledge Management (CIKM'03)*, (pp. 504–507), New Orleans, LA.

White, R. W. and Jose, J. M. (2004). An Implicit System for Predicting Interests. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'04)*, (p. 595), Sheffield, UK.

CHAPTER 10

EDUCATIONAL KNOWLEDGE DOMAIN VISUALIZATIONS: TOOLS TO NAVIGATE, UNDERSTAND, AND INTERNALIZE THE STRUCTURE OF SCHOLARLY KNOWLEDGE AND EXPERTISE

PETER A. HOOK AND KATY BÖRNER
Indiana University

> [M]ost cognitive scientists believe, learning best begins with a big picture, a schema, a holistic cognitive structure, which should be included in the lesson material—often in the text. If a big picture resides in the text, the designers' task becomes one of emphasizing it. If this big picture does not exist, the designers' task is to develop a big picture and emphasize it[.] (West, Farmer and Wolff, 1991, p. 58).

1. INTRODUCTION

Today, we attempt to access all humanity's knowledge and expertise using search engines such as Google. This works well for fact retrieval. However, search engines do not enlighten the user as to the inherent structure of the information being searched or give the user feedback as to its completeness. There is no 'up' button. The user is not able to see what dataset was queried, how the entries in a search result set relate to each other or how the retrieved entities relate to the entities that were not retrieved. Effective approaches to information access and management need to take into account the human user's perceptual and cognitive capabilities. Humanity is in true need of better tools to filter, navigate, understand, and utilize (scholarly) knowledge.

This chapter discusses domain maps as an alternative means to organize, navigate, and internalize scholarly knowledge. We first discuss the educational uses of maps and the benefits of information visualization and spatialization for education. Subsequently, we introduce thematic maps, cognitive and concept maps, knowledge domain visualizations, and information spaces employing the metro map metaphor. All four are visual representations of geographic or abstract semantic spaces. Given that our interest is in the access, management, and internalization of scholarly knowledge, knowledge domain visualizations are discussed at greater length. To this end, we discuss how the educational use of knowledge domain visualizations is supported by the semantic network theory of learning. We also discuss some of the elements of good knowledge domain map design. These are drawn from visual perception principles and the study of human memory, and cognition. The final section projects a potential future of educational knowledge domain visualizations.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 187–208.

2. EDUCATIONAL USAGE OF MAPS

This section discusses the utilization of spatial learning strategies, big picture views, and conceptual maps in educational settings. There has been a long history of spatial learning strategies in the field of education (Holley and Dansereau, 1984). With the development of the semantic network theory of learning, many educational theorists began creating and implementing spatial learning techniques. These spatial representations of knowledge are used as: (1) learning tools, (2) evaluation tools, (3) curriculum and instruction planning tools (via both macro and micro maps) and (4) tools to facilitate cooperative learning (Milam, Santo, and Heaton, 2000) and cooperative scientific research or collaboratories (MacEachren, Gahegan, and Pike, 2004).

In education, visual representations of the big picture view are applied for diverse reasons. First, they provide a structure or scaffolding that students may use to organize the details of a particular subject. In this fashion, information is better assimilated with the student's existing knowledge and the visualization enhances recall. Second, big picture displays make explicit the connections between conceptual subparts and how they are related to the whole. Third, big picture representations help to signal to the student which concepts are most important for them to learn (West, Farmer, and Wolff, 1991).

One commonly used instantiation of big picture views are concept maps. (See Figure 1 and also the subsequent section on Cognitive and Concept Maps.) A concept map is made up of four core elements: (1) shapes or nodes—representing core elements of a concept, (2) connectors or links between the shapes or nodes, (3) connecting

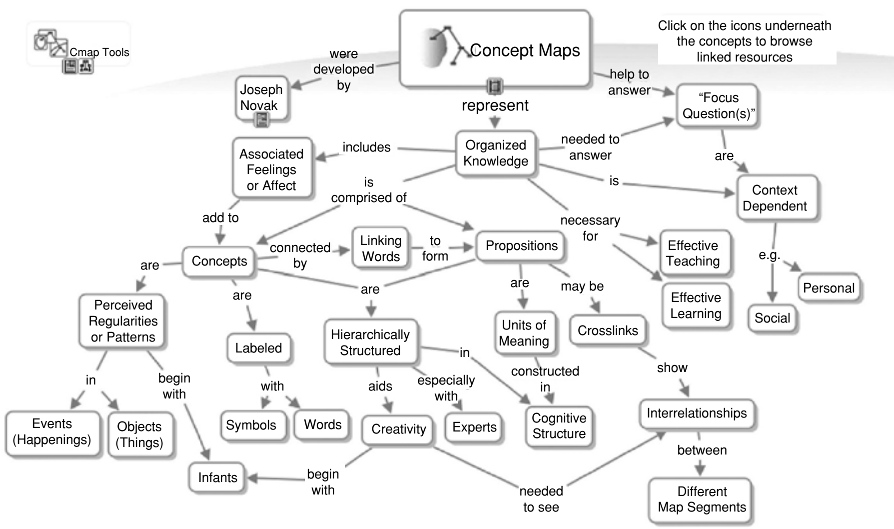

Figure 1: Exemplary concept map showing information about concept maps. Created by Joseph Novak and rendered with CMapTools. Copyright: Institute for human and machine cognition (IHMC). Used with permission.

words—that describe how two nodes are related, and (4) patterns—such as a hierarchical or circular ordering of the nodes (Milam, Santo, and Heaton, 2000). There are at least nine different categories of connecting words between nodes: subsuming, similarity, quantity, enabling, causal, timing, dissimilarity, equivalence, and categorizing (West, Framer, and Wolff, 1991). Concept maps have also been referred to as mind maps, pattern notes, brain patterns, spider maps, networks, semantic maps, semantic networks, and semantic webs (predating and different from the WWW Consortium's creation and promotion of the Semantic Web for the Internet) (Milam, Santo, and Heaton, 2000).

Concept maps are used to evaluate a learner's understanding of a concept by having the learner display the key elements of the concept in a spatially structured, interconnected layout. This allows an instructor to detect fundamental inaccuracies of a student's conceptual schema for a particular piece of knowledge and to take steps to remedy the inaccurate schema. As teaching tools, the concept maps of the teacher or other experts are presented to the student to use as cognitive scaffolding for assimilation of the topic being studied. Presently, it is common to find un-labeled cognitive map worksheets (still containing blank nodes and connecting lines), for students to fill in as they study a particular topic. These worksheets are contained in the preprinted workbook that accompanies many grade school textbooks (Feather, Snyder, and Hesser, 1993). Furthermore, a variety of tools exist that ease the creation of concept maps.[1]

The study of inherently text based subjects may be significantly enhanced by the benefits of information visualization and spatialization.[2] A concept or domain map responds to the human brain's preference for information that is organized spatially. Information is easier to remember if it can be stored and located spatially (Miller, 1968; Winn, 1994). In fact, the spatial dimension of memory and learning is more important than non-spatial features such as color, shape, or action (Miller, 1968). Additionally, spatial information is easily committed to memory (Allan, 1999).

A domain map of a subject adds a spatial component to a topic that is unavailable in a strictly linear presentation such as a table of contents appearing at the beginning of a textbook or a list of navigable topics in a database. Users of a domain map will have increased recall of the topic by having their memories stimulated by the spatial layout of the domain map. In part, this is due to the fact that recognition is faster and more accurate than recall. "[V]isual memory traces of objects and scenes are stored as part of the processing mechanism; thus it is not necessary for an object to be fully processed for recognition to take place" (Ware, 2004, p. 299). Additionally, humans store textual and visual information in different areas of the brain. This is known as the 'dual coding' theory of memory (Kulhavey and Stock, 1994; Ware, 2004).

The visual cues inherent in a domain map that the user stores in his or her nonverbal (non-textual) memory region should be highly effective in cuing memory stored in the user's verbal (textual) region:

---

[1] See: CMap—http://cmap.ihmc.us/; Inspiration http://www.inspiration.com/; Mindmanager http://www.mindjet.com/us/; SMARTIdeas http://www2.smarttech.com/st/en-US/Products/SMART+Ideas/; Visimap—http://www.visimap.com/

[2] Spatialization is the spatial representation of non-spatial data.

> "Intact map images retain the spatial characteristics of the objective stimuli, and the information within them becomes simultaneously available when they are brought into working memory. Such images have an advantage as far as the cuing of text retrieval is concerned. The advantage derives from the fact that attention can be shifted from location to location across the map image without exceeding the limits of working memory" (Kulhavy and Stock, 1994, p. 155).

Kulhavey and Stock (1994) ran a series of experiments to test if the associative recall of textual information stimulated by maps was a result of structural features (the spatial layout of the map) and not just non-spatial features such as landmarks, labels, drawings, topographic symbols, and the shape, size and color of depicted items. They concluded that the spatial arrangement of map content was a crucial element in cuing the textual information.

Newbern, Dansereau, and Patterson (1997) confirmed Kulhavy and Stock's co-joining of spatial and verbal memory hypothesis. The authors found that students learning a concept from a knowledge map (in this case a text heavy flow diagram of biological processes) had better recall of the concept than students who learned the same concept from a conventional textual write-up. Furthermore, students were better able to pinpoint where information was located on the knowledge map than with the textual write-up of the subject.

Furthermore, remembering that something is located in a specific place, and what that something is are two different memory tasks. However, they complement each other. When object and location become linked, they serve as memory cues for one another. The object will trigger recall of the spatial location, and the spatial location will trigger recall of the object (Allen, 1999). For all of these reasons, domain maps introduce important spatial and visual elements to subjects that are most frequently presented in a text-heavy manner. It is these spatial and visual elements that amplify cognition and enhance learning. In sum, spatial representations of knowledge appear to improve access to, retrieval, and management of knowledge. Research findings that point to the educational use of spatial learning strategies, big picture views, and conceptual maps should help improve the design and utilization of educational knowledge domain visualizations explained in the next sections.

3. SPATIAL/VISUAL REPRESENTATIONS OF KNOWLEDGE

Cartography has long used spatial representations of the planet as a substrate for additional variables. Maps imposing one or more variables onto a literal spatial substrate are referred to as thematic maps. Spatial representations of knowledge have also been used in the fields of wayfinding, psychology, and education to model an individual's understanding of particular concepts. These are known as cognitive maps and concept maps. More recently, the field of information science has begun to represent entire domains of knowledge using spatial representation techniques. These are known as knowledge domain visualizations (KDVs) (Börner, Chen, and Boyack, 2003).

4. THEMATIC MAPS

Cartography is the science dealing with the representation and transformation of spatial information (MacEachren, 1995; Skupin, 2004). There are two major types of maps: (1) general purpose or reference maps which display numerous features of a landscape and emphasize location, and (2) thematic maps, which display the geographical characteristics of select statistical phenomena and often focus on a single theme (Slocum, 1998). Thematic maps are designed to convey the structural characteristics of geographic distributions (Chen, 2003). "[T]hey attempt to show the spatial distribution of one or a few variables on the earth's surface, variables that may not be directly perceptible in the environment at all (e.g., disease rates)" (Montello et al., 2004, p. 255). Thus, the major design elements of thematic maps are a geographical, or spatial, substrate and an overlay of additional information. As will be discussed later, these same elements are present in most knowledge domain visualizations.

Figure 2 shows a typical thematic map. It depicts the results of the 2004 United States presidential election. The size of the less populous states won by George W. Bush might give the impression that the election was a lopsided victory in Bush's favor. Figure 3 shows a cartogram of the same election results. A cartogram is a thematic map in which the spatial substrate has been intentionally distorted to better reveal the characteristics of the information overlay (Slocum, 1998). In this case, the cartogram in Figure 3 provides a better indication of the competitiveness of the 2004 presidential

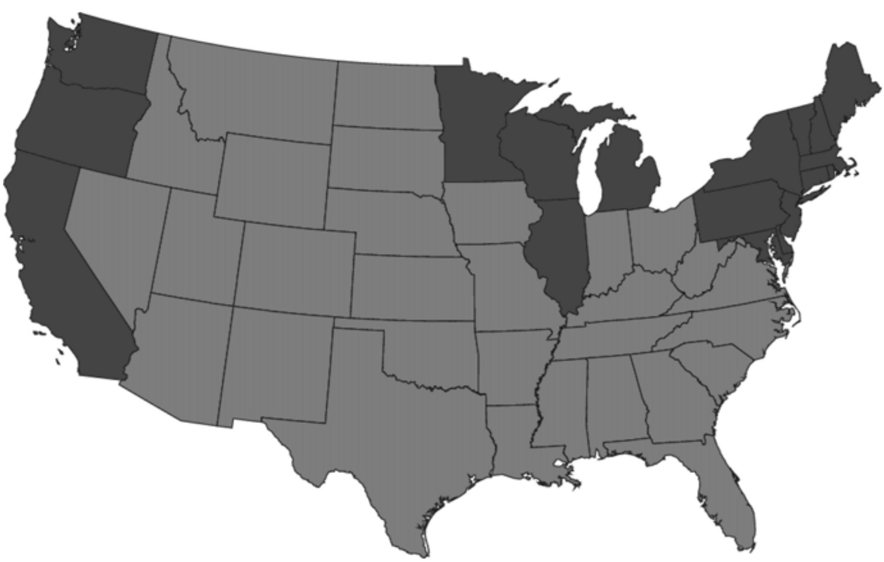

Figure 2: A typical thematic map displaying the results of the 2004 presidential election. Red States won by George W. Bush. Blue States won by John Kerry. Created by: M. Gastner, C. Shalizi, and M. Newman. Permission to reproduce given on Website: http://www.cscs.umich.edu/∼crshalizi/election/

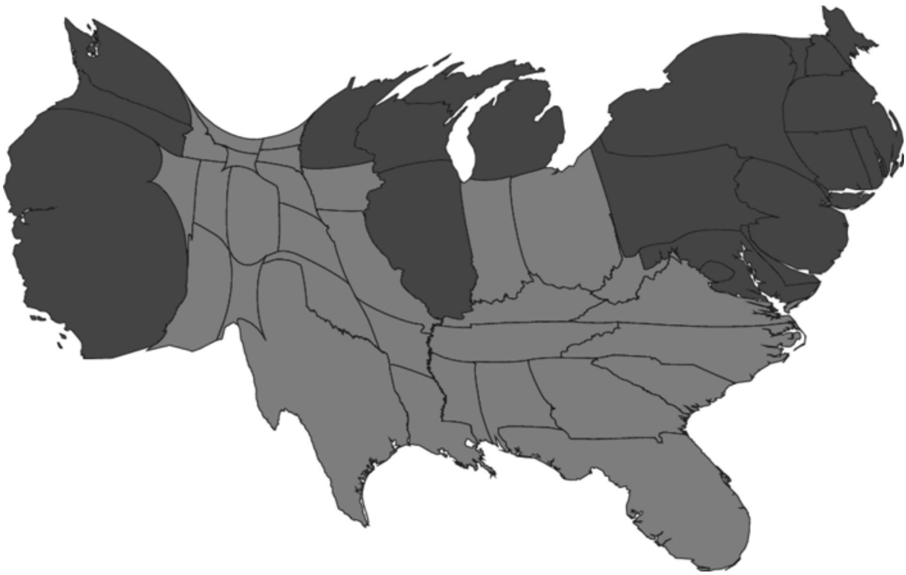

Figure 3: A Cartogram of the 2004 presidential election with the geo-spatial substrate distorted to reflect the population of each State. Red States won by George W. Bush. Blue States won by John Kerry. Created by: M. Gastner, C. Shalizi, and M. Newman. Permission to reproduce given on Website: http://www.cscs.umich.edu/~crshalizi/election/.

race. It reflects the relative closeness of both the popular vote and the vote in the Electoral College. Such manipulation of the underlying spatial substrate to showcase desired relationships is also common in the production of domain maps.

5. COGNITIVE MAPS AND CONCEPT MAPS

Cognitive maps refer to the representation of spatial information inside the mind of a particular subject (Golledge, 1999). They are used in wayfinding, the science of how people navigate spatial environments. Cognitive maps used in wayfinding employ most of the graphic elements of concept maps used in the field of education. (1) Shapes or nodes are referred to as points. In wayfinding, points can be either landmarks or reference nodes. (2) Connectors or links between the shapes or nodes are lines (routes, paths, or tracks). (3) Connecting words used in educational concept maps do not have an exact parallel in wayfinding or cartography. In wayfinding and cartography, the connecting relationships between nodes are usually implicit in the spatial arrangement of the map or are labels given to routes between nodes. The latter case includes such information as the name and surface features of the route. However, the labels almost never explicitly state notions of causation or other relationships. (4) Patterns are also present in the cognitive maps of wayfinding. Patterns are conceptualized as "regions, neighborhoods, and topological containment or inclusion," and these patterns are also detectable in three dimensional notions such as density and changes in elevation (Golledge, 1999,

p. 15). "Places and locations form spatial distributions; tracks, paths, and roadways form networks; landmarks and nodes from hierarchies. All of these combine to represent the total knowledge structure (cognitive map)." (Gollege, 1999, p. 20).

As introduced previously, concept maps are a representational tool used in the fields of education and psychology. They represent non-spatial, abstract concepts using spatial representations, see Figures 1 and 4. In wayfinding, the user's internal cognitive map models actual spatial reality. Education and psychology use the spatially arranged, networked nodes of concept maps to represent ideas or processes that are not inherently manifest in three dimensional space. While there are subtle differences between cognitive maps, concept maps, and all of the other similar types of maps used in the fields of education, psychology and wayfinding, they will henceforth be collectively referred to as 'concept maps'—internal spatial representations of external reality or ideas.

Figure 4 portrays a concept map with an underlying substrate that was created with procedural rigor. It reflects the consensus of the concept maps created by 13 graduate students in an education course in which the subject matter was integrating technology into the classroom (Kealy, 2001). It is unique in that it was created using the input of an entire class. It is typical in that it has all four elements common to concept maps

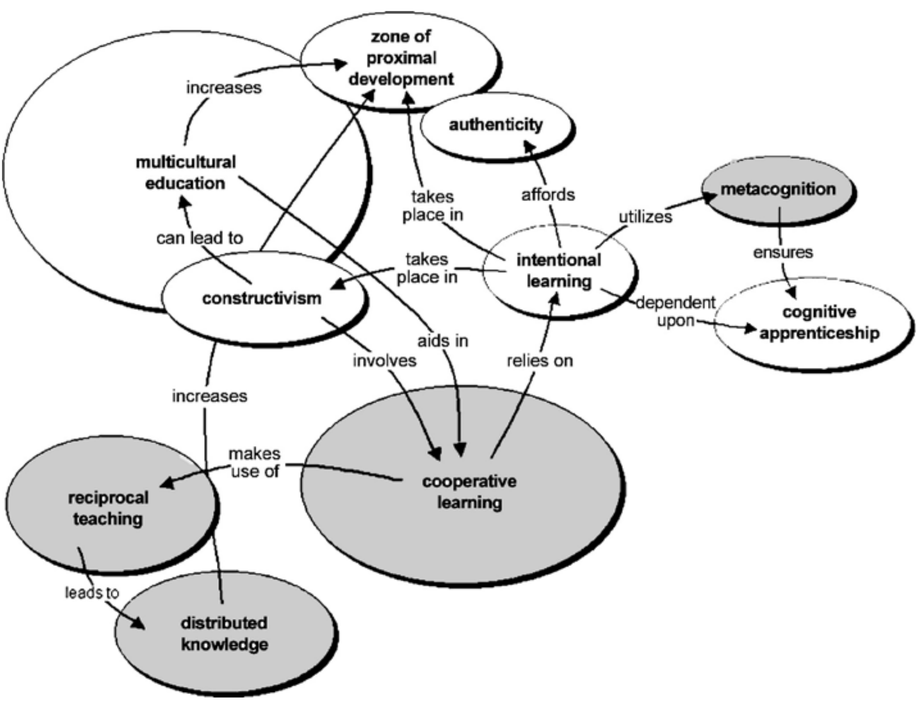

Figure 4: Exemplary concept map showing the subject matter of a graduate level education course on integrating technology into the classroom (Kealy, 2001, p. 345.) Copyright: Baywood Publishing Co., Inc. Used with permission.

used in the field of education: (1) nodes, (2) links, (3) connecting words that describe how the nodes are related, and (4) patterns (in this case a spatial ordering of the nodes suggesting their semantic proximity.)

The concept map in Figure 4 is one of the very few known to the authors from the field of education that employs rigorous methods to derive both the spatial layout of the map and the characteristics of the nodes. (In contrast, such production techniques are commonly used in the creation of knowledge domain visualizations (see below)). The spatial layout, or substrate, of the concept map in Figure 4 was created using multidimensional scaling ('MDS') techniques (see generally Kruskal and Wish (1984)). These techniques also revealed that the two most important dimensions for the concept map were importance (important vs. unimportant) and utility (theoretical vs. practical). Surveys of the class participants led to the creation of the nodes with size representing importance and hue representing utility.

6. KNOWLEDGE DOMAIN VISUALIZATIONS

Knowledge domain visualizations (KDVs) (Börner, Chen, and Boyack, 2003) are the graphic rendering of bibliometric data designed to provide a global view of a particular domain, the structural details of a domain, the salient characteristics of a domain (its dynamics, most cited authors or papers, bursting concepts, etc.) or all three, see Figures 5 and 6. KDVs are also referred to as domain maps and the process of their creation as domain mapping.

Research on KDVs is conducted in the fields of information science (and its related sub-disciplines of scientometrics and bibliometrics) (Bernal, 1939; Börner, Chen, and Boyack, 2003; Braam, Moed, and Van Raan, 1991a,b; Callon, Law, and Rip, 1986; De Solla Price, 1965; Garfield et al., 1964; ISI, 1981; Marshakova, 1973; McCain, 1990; Small, 1973, 1999; White and McCain, 1998). Information science investigates the origination, dissemination, collection, organization, storage, retrieval, interpretation, and use of information (Rubin, 1998). Scientometrics aims at the quantitative study of science. Bibliometrics studies scholarly production based on bibliographic data (Buter, Noyons, and Van Raan, 2004). Consequently, KDV's generated in these three different areas serve different information needs and have a different appearance.

Today, KDVs are typically generated by KDV experts in close collaboration with domain experts using a rather time consuming, laborious process that requires extensive knowledge in terms of data sampling, data analysis, layout, interaction design and last but not least, the interpretation of the resulting maps. Few companies and institutions have the resources to pay for customized maps that reveal the key information (major experts, papers, inventions, emergence of new research frontiers) for their area of interest. This is unfortunate as such knowledge domain visualizations can help to analyze research productivity and lead to more informed decisions as to the allocation of scarce resources.

KDVs are seldom used in educational settings today. This is most likely due to the expense of their generation. In addition, many KDVs are hard to read for non-experts and frequently need to be translated into the jargon of the domain to be truly useful. This is unfortunate in light of the compelling evidence of the utility of maps stemming from

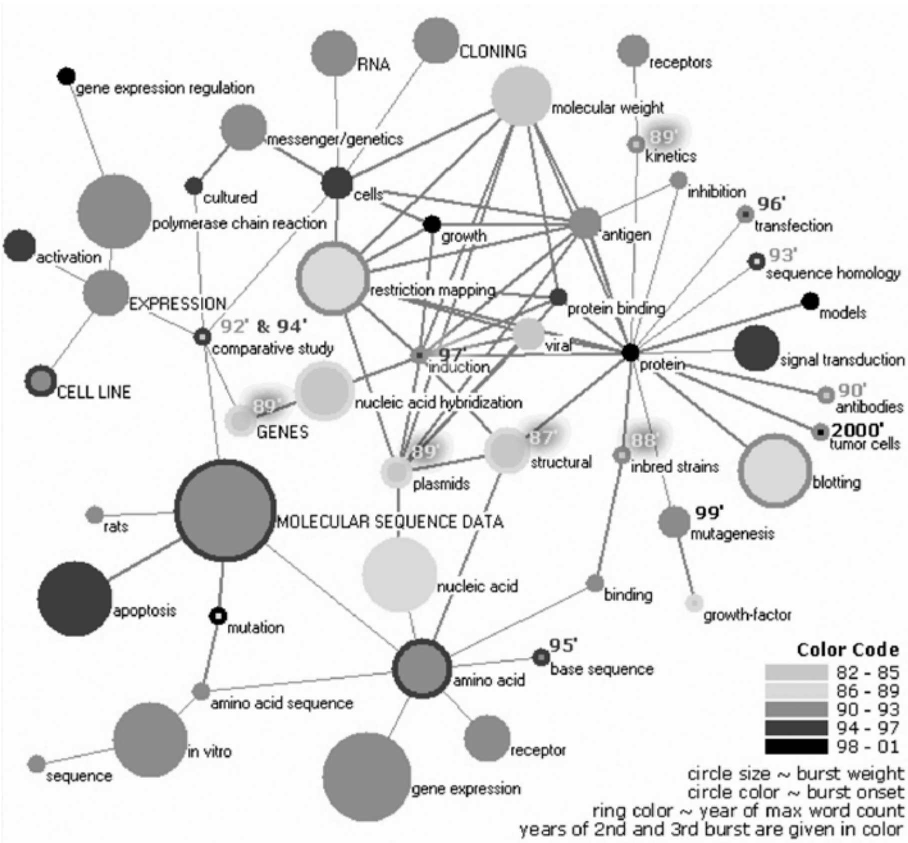

Figure 5: Exemplary knowledge domain visualization showing a node link diagram of keywords appearing in PNAS between 1982 and 2001. It portrays the 50 most frequent and bursty words used in the top 10% most highly cited PNAS publications (Mane and Börner, 2004). Copyright: The National Academy of Sciences of the USA. Used with permission.

the fields of cognitive science, educational psychology, cartography, citation analysis, scientometrics, complex systems, network studies, and bibliometrics. Well designed KDVs have the ability to facilitate understanding, recall, and to convey to the user the schematic, geo-spatial, temporal, semantic, or social organization of the underlying domain. KDVs differ from concept maps currently used in the field of education in the scope of the domain being represented. In education, the maps are micro thematic—most often dealing with a single theme or concept within a field. In contrast, information scientists often create large scale knowledge domains focusing on entire fields or all of science.

Also, the educational community places more emphasis on the importance of drawing and labeling the connections between nodes with action phrases such as “gives rise to,” and “causes.” These explicitly labeled connections, linkages, or notion of causality

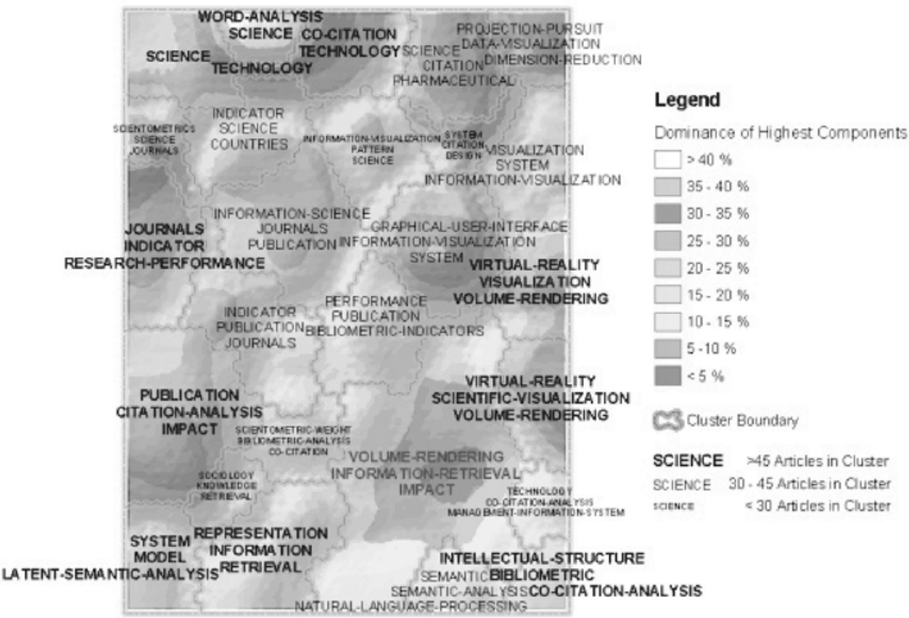

Figure 6: Another exemplary knowledge domain visualization showing a geospatial rendering of semantic space relevant to articles related to knowledge domain visualizations (Börner, Chen, and Boyack, 2003). Created by André Skupin. Copyright: American Society for Information Science and Technology. Used with permission.

are largely missing from KDVs. While suggested implicitly by the spatial proximity of related topics, such issues as causality would be hard to make explicit using automated techniques currently used by information scientists. Garfield's Histograms (Garfield et al., 1964) of the discovery of the structure of DNA have these explicit, temporally labeled connections. However, the use of labeled connections does not rise to the level of use in concept maps used in education.

7. METRO MAP METAPHOR

Henry Beck's (1933) redesign of the London Underground (subway) map was a revolutionary advancement in graphic design that has recently begun to be applied to non-geographic, information spaces (Nesbitt, 2004). Beck did three things. (1) He removed most of the surface features from the map, thereby reducing clutter. (2) He decoupled the map from the need to show accurate scale distances between all stations. In other words, he distorted the spatial substrate to give the center, denser portion of the map more area so that its features could be disambiguated and made clearer to the viewer. Also, stations were portrayed at equal distances from one another, when in reality this was not the case. (3) Finally, he unraveled the sinuous nature of the map

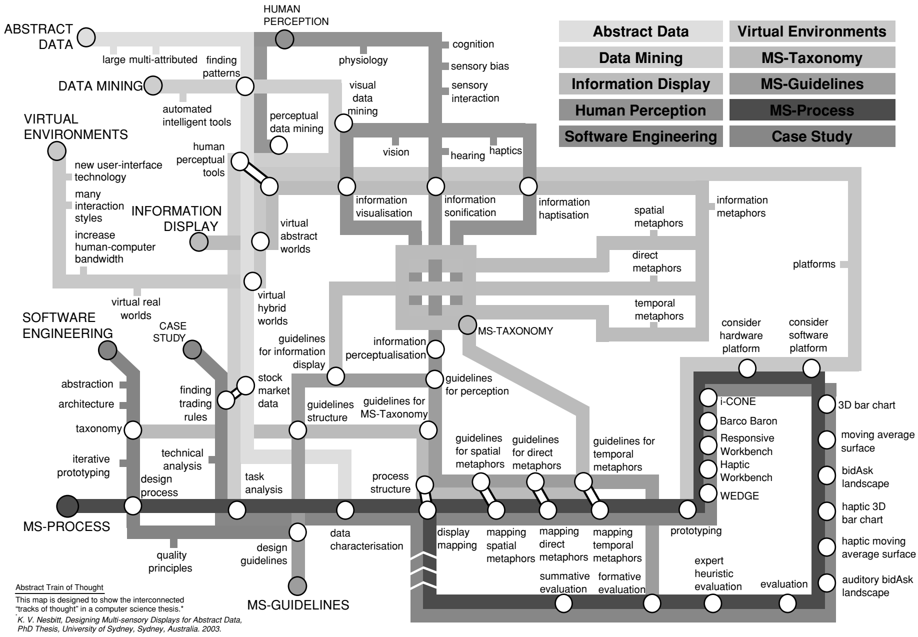

Figure 7: Exemplary knowledge domain visualization utilizing the metro map metaphor. It shows the interconnecting lines and nodes of thought for Keith Nesbitt's doctoral dissertation (Nesbitt, 2004). Copyright: IEEE. Used with permission.

and made all routes orthogonal with only forty-five or ninety degree angles (with tiny rounded elbows) (Garland, 1994). These easy to comprehend features soon became common on metro maps used throughout the world, see (Ovenden, 2003).

Figure 7 shows the metro map metaphor applied to a non-geographic, information space. It portrays the interconnecting lines and nodes of thought for Keith Nesbitt's doctoral dissertation. The metro map metaphor has the potential to merge domain maps used in the field of education (concept maps) with those coming out of the field of information science (knowledge domain visualizations).

The metro map metaphor contains most of the four elements of concept maps: (1) nodes—representing core elements of a concept (the stations), (2) connectors or links between the shapes or nodes (the route lines) and (3) patterns—such as a hierarchical or circular ordering of the nodes. The missing element, explicit labels describing the connections between nodes, may also be added. Already, the route lines demark subject groupings and their interconnections. In fact, the metro map metaphor allows for more complex, intertwined interactions among subject groupings than is possible with traditional groupings of bounded regions. Additionally, the layout of the nodes, route lines, and perhaps even clearly demarcated spatial regions may be rigorously performed using the same information science techniques that produce knowledge domain visualizations.

Table 1: Comparison of thematic maps, concept maps and knowledge domain visualizations

	Substrate			
Map type	Usage of nodes & edges	Rigorous spatial layout	Labeled connections (causality, equivalence, similarity, etc.)	Overlay
Thematic Maps	rarely	yes	no	yes
Concept Maps	yes	very rare	yes	yes
Metro Map Information Spaces	yes	potentially	no	yes
Knowledge Domain Visualizations	yes	yes	rarely	yes

8. SUMMARY OF SUBSTRATE/OVERLAY INFORMATION SPACES

Table 1 shows how KDV's compare with thematic maps, concept maps, and information spaces employing the metro map metaphor. All four map types are composed of two parts: a geographic or base map, and a thematic overlay. Thematic maps and KDV's use numerous techniques to structure the spatial component of the domain map. While concept maps have a spatial substrate, they are seldom created with rigorous methods. While none currently exist that we are aware of, metro map information spaces have the potential to be created with rigor as to the spatial substrate. Also, while thematic maps typically use a rigorously ordered spatial substrate, they rarely use node-link diagrams. On the other hand, nodes and links are a crucial component of most KDVs, and all concept maps and metro map information spaces. On top of the spatial substrate employed by all four types of maps are placed the thematic components of the map. In the case of knowledge domain visualizations these might be topic words, authors, specific works, contributing institutions, etc.

Recent efforts aim to create tools that will ease the generation of highly usable maps of the sciences (Borner, Chen, and Boyack, 2003). We believe that KDV's can benefit enormously from user studies and case studies conducted for other types of maps.

9. KDVs SUPPORTED BY THE SEMANTIC NETWORK THEORY OF LEARNING

According to the **semantic network theory of learning**, human memory is organized into networks consisting of interlinked nodes. Nodes are concepts or individual words. The interlinking of nodes forms knowledge structures or schemas. Learning is the process of building new knowledge structures by acquiring new nodes. These new nodes are interrelated with existing nodes and with each other. When learners form links between new and existing knowledge, the new knowledge is integrated and comprehended (Jonassen, Veissner, and Yacci, 1993). In other words, "[w]e learn new concepts by associating them with familiar ones" (Chen, 2003, p. 69).

Learning is the process of reconciling new phenomenon with existing frameworks of understanding. If reconciliation is not possible, then the existing framework(s) must be modified to accommodate the new knowledge. KDVs are an effort to explicitly convey the underlying structure of a domain to the user so the user can internalize the framework presented in the KDV and reconcile it with his or her existing framework. A good understanding of the structural organization of a domain is a better predictor of being able to problem solve in that domain than aptitude as measured by standardized test scores (Jonassen, Veissner, and Yacci, 1993). Research has also shown that what separates expert and novice problem solvers is the well-developed and interconnected knowledge networks of the experts that facilitate both the interpretation and solution of the problem (Jonassen, Veissner, and Yacci, 1993). KDVs should assist in providing the structural knowledge of the domain that will allow novices to become better problem solvers.

Any **expert** attempting to convey information will instinctively structure the information based on his or her knowledge schema for that particular domain. Readers who are able to discern the author's schema are better able to remember and internalize the information than those who do not recognize the author's organizational framework. Studies also show that learners who are regarded as having good reading comprehension instinctively search out the author's organizational structure and use it to recall content (Jonassen, Veissner, and Yacci, 1993). Experts in a domain have fewer problems making spatial representations of the domain than non-experts (Breuker, 1984). It is thought that experts possess structurally unifying notions of the domain. It is these notions that can be brought out in a domain map and conveyed to a user.

A **novice user** benefits from a domain map as it expressly conveys the organizational structure of the domain. The user does not have to piece it together from such things as chapter headings, subheadings, and non-content signal words in the text that convey the relationship between ideas (for example, therefore, next, in contrast, on the other hand, etc.). Domain maps provide the user with the necessary visual scaffolding to assimilate and comprehend the domain. A structural understanding of a domain is also the basis of a deeper, more conceptual, and abstract understanding of a domain (Jonassen, Veissner, and Yacci, 1993). Furthermore, structured items are better recalled than unstructured items (Jonassen, Veissner, and Yacci, 1993).

Hopefully, KDVs will become as central to learning as the periodic table of the elements is to chemistry, planetary maps are to geographers, star charts are to astronomers, and city maps are to out-of-town visitors. Like the periodic table, KDVs provide the scaffolding upon which a learner may organize and build new understanding. Analogous to the periodic table, a KDV's transmission of the underlying structure of the domain should also have predictive qualities. By seeing where a topic is placed on a KDV, the user may draw from his or her store of existing knowledge about adjacent topics to begin to understand what an unknown topic is about.

Additionally, much has been written about the existence of multiple learner types (visual, aural, tactile, etc.) (Hook, 2002). Most likely, KDVs will be most effective with visual learners. However, even if KDVs appeal more strongly to visual learners, it is good to add another beneficial tool to the overall package of teaching tools so that collectively they may be used to reach all learners.

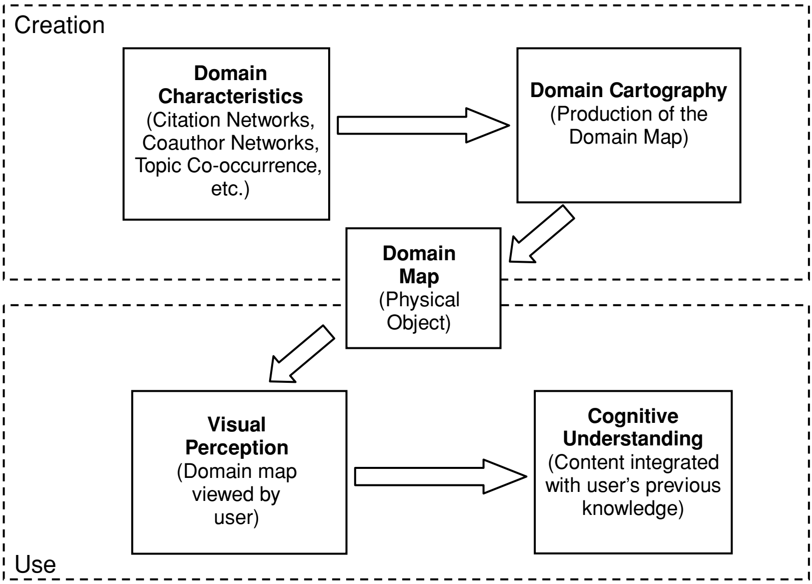

Figure 8: The process of KDV creation and usage

10. THE DESIGN OF EFFECTIVE KNOWLEDGE DOMAIN VISUALIZATIONS

The design of effective KDVs should be based on a deep understanding of visual perception principles and human cognitive abilities. The utilization of KDVs in educational settings requires knowledge of how maps can be used to store, access, manage, and communicate information. Figure 8 sketches the process of domain map generation and usage. Map generation can be seen as a representational process aimed to convey the structure and dynamics of a knowledge domain.

The knowledge domain exists independent of the observer. Bibliometric data on the domain is harvested, cleaned, analyzed and displayed by the KDV expert. Once the domain map is created, it is presented to the viewer (typically a domain expert) for interpretation. The viewer first perceives the map using low level visual perception (Palmer, 1999; Ware, 2004). Next, the viewer employs higher level cognitive processes to internalize and understand the domain map by reconciling the new content with the viewer's previous knowledge.

Numerous articles detail the process of creating KDVs (Börner, Chen, and Boyack, 2003; Chen, 2003). Subsequently, we discuss some of the perceptual and cognitive issues involving KDVs. Our ultimate goal is the identification of visual perception and cognitive principles that inform the generation of KDVs that truly amplify cognition.

It is interesting to note that the validation of large scale maps of science is a serious problem. Domain *experts* have a very specialized and often subjective view of their area of expertise. Confronted with a map of all of biology they need to zoom 10 to 15 times (depending on their level of expertise) until they recognize papers or people's

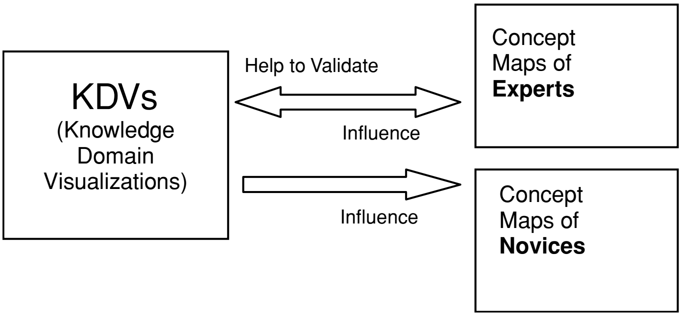

Figure 9: Interaction between KDVs and concept maps of experts and novices

names with which they are familiar. We believe KDVs can play an important role in enhancing and fleshing out the concept maps of expert users by reminding them of implicit knowledge and pointing out related research areas. Also, KDVs provide a cognitive structure that will influence and potentially rearrange a *novice* user's internal representation of the domain, see Figure 9.

Subsequently, we review major visual perception principles and human cognitive abilities that might help guide the design of effective KDVs and their utilization for educational purposes.

11. VISUAL PERCEPTION PRINCIPLES

A number of excellent textbooks exist on visual perception principles (Palmer, 1999; Ware, 2004). Here we discuss principles that are directly related to the design of highly informative and readable KDVs. Several of them are related to the perception of space and the grouping of objects in space.

Spatial proximity as an expression of semantic closeness is one of, if not the central, metaphors used for the design of KDVs. User studies have shown that subjects are able to equate distance with similarity. This is also known as the "distance-similarity metaphor" (Montello et al., 2003). This ability appears to be independent of the user's background or familiarity with spatial data. In a series of experiments, Skupin and Fabrikant (2003) have shown that users are able to associate "(1) interpoint distance with the concept of document similarity in a document collection; (2) graphic clusters representing the information content and structure of a digital collection with concentration of related documents; and (3) graphical change in resolution (zoom-in) with different levels of detail in a document collection (hierarchical order)" (p. 110). Spatial proximity and grouping is **processed preattentively** and hence rather fast (Ware, 2004).

Optical illusions. The perception of distances can be distorted by **vertical illusion** effects. Montello et al. (2003) conducted studies in the context of document spaces and showed that for two items of equal distance from a referent third item, the item oriented along the vertical access from the referent item will be perceived as further away than the item oriented along the horizontal access from the referent item. Furthermore, the perception of spatial distance is affected by intervening items that lie between and adjacent to the items being compared for proximity. These are known as **emergent feature effects** (Montello et al., 2003). For instance, once three or more items lie in a linear manner between Item A and referent Item C, Item A will be perceived as more proximate to the referent Item C than Item B even though Item B is closer in distance but has no intervening linear points.

Clustering is another visual perception phenomenon that impacts the perception of proximity. Given two items A and B, Item A will be perceived as closer to a referent Item C, if A and C are in the same cluster. This is true even if Item B is actually much closer to Item C, but not in the same cluster as A and C.

Gestalt principles can be employed to further support the perception of semantically related articles, authors, journals, keywords, domains of sciences, etc. See generally (Koehler, 1947; Moore and Fitz, 1993). For example, items that have a similar appearance tend to be perceived as a unit. For this reason, different types of objects depicted in a KDV should be distinguished by different shapes. It should be noted that items with the same shape are more readily perceived as belonging to the same category than items of the same color (Ware, 2004).

Connectedness. Connections can be "a more powerful grouping principle than proximity, color, size or shape" (Ware, 2004, p. 191). Linear connections (links) and separations (boundaries) help to offset the unwanted perceptual effects set out above. Also, they can help with overcoming the inevitable distortions that arise in the automated conversion of high dimensional data into a low dimensional spatial layout (Skupin and Fabrikant, 2003). The educational psychology community has long felt that connecting lines are a crucial element in representing knowledge to a viewer. While difficult to derive using automated means, connecting elements are a quick and powerful way to convey information about a domain. Whenever possible, they should be included in a KDV. This is true even if they have to be added by hand based on the knowledge of domain experts. The addition of connecting elements derived from expert observation and not from bibliometric data is consistent with Buter's appeal for the use of both quantitative and qualitative bibliographic maps in understanding knowledge domains (Buter, Noyons, and Van Raan, 2004).

Visual or Organizational Hierarchy. Hierarchies are a crucial component of how humans process, remember, and utilize information. Likewise, the hierarchical structuring of information on maps is equally important. Not every piece of information on a map is of equal significance. Graphic techniques are able to signal the most important concepts of a map, or those elements that should be visually processed first. The most important items should be rendered with the greatest contrast to their surroundings. Less important items should be rendered with less contrast (Chen, 2003). Additionally, all seven of Bertin's visual variables (position, shape, orientation, color, texture, value and size) may be used to reveal the data hierarchy of the knowledge domain visualization (Bertin, 1983). Label sizes (letter size, the use of capital and lower case letters, and the size of the map symbols) are good tools to convey the visual hierarchy. Graduated color choices and differing line widths demarking boundaries are another way to convey the hierarchical structure of a map (Skupin, 2004).

Distinguishable Regions or Clusters. Studies show that good map learners divide a map into regions and focus their attention on the regions that they have yet to understand

and internalize (Allen, 1999; Thorndyke, 1981; Thorndyke and Hayes-Roth, 1982). Consequently, different regions of the map should be well bounded, easily demarcated, readily distinguishable, and memorable.

Landmarks. Both cartography and the field of wayfinding have established the importance of landmarks in understanding maps. Landmarks help lighten the cognitive load on the user. They help orientate the user while viewing/ experiencing the spatial representation of a map. "Landmarks are often noticed and remembered because of dominance of visible form, peculiarity of shape or structure, or because of sociocultural significance" (Golledge, 1999, p. 17).

The significance of a landmark may be either objective or subjective. Some objects are identified as landmarks by most members of a group. Other landmarks have significance because of individual experiences with the object. Landmarks are analogous to important nodes in the schematic network theory of learning with which new concepts are associated. Landmarks provide anchor points from which the rest of the information on the map is organized. Landmarks act as "primary organizing features in cognitive maps by dominating a spatial classification or clustering process to facilitate environmental knowing and understanding" (Golledge, 1999, p. 17).

Major landmarks on domain maps should be readily distinguishable from other nodes. It is not enough that they be labeled with bolder text, represented by larger dots, extend higher into space, or otherwise have a larger shape. They should have an idiosyncratic component. They should be unique and preferably represented by an image. The rationale for this may be drawn from Winn and Holliday's principle number four for the design of diagrams and charts: "Include small pictures in diagrams to teach concept identification, especially with students of low verbal ability" (Winn and Holliday, 1982, p. 286). Visually prominent landmarks will also facilitate recall by being easy to find again. They will also serve as a mnemonic for better remembering the information on the map.

Colors. Color is often the best variable to use to code data categories (Ware, 2004). Color coding makes it easy for a viewer to categorize different objects (Ware, 2004), and presumably different concepts. Color may be used to differentiate bounded regions and contribute to the overall legibility of a KDV. Color choices also contribute to the viewer's perception of the information hierarchy of the KDV.

Ware (2004) lists a set of criteria to consider when using color to encode data. (1) Colors should be distinct. (2) Unique hues should be used (red, green, yellow, blue, black, and white.) (3) Colors should contrast with the background. (4) Color blindness should be taken into account. (5) Only about five to ten color codes may be rapidly distinguished. (6) Larger color coded areas make it easier to perceive contrasts in color coding. Extremely small color-coded areas should not be used as it becomes hard to distinguish between the colors. Furthermore, "[w]hen large areas of color coding are used, for example with map regions, the colors should be of low saturation and differ only slightly from one another. This enables small, vivid color-coded targets to be perceived against background regions" (Ware, 2004, p. 125). Finally, (7) the mapmaker should consider cultural color coding conventions (red = hot, blue = cold).

Ware recommends the use of the following twelve colors to encode data: red, green, yellow, blue, black, white, pink, cyan, gray, orange, brown, and purple. The first six are to be used before the second six (Ware, 2004). The use of black or white borders around colored symbols helps to ensure that the symbols stand out from the background colors (Ware, 2004). Additionally, colors may encode more than one variable. For instance, the hue of a color may represent one variable while the saturation of a color represents a second variable. However, such maps are extremely hard to read and the second variable should be represented using an additional information channel such as texture (Ware, 2004).

12. CARTOGRAPHIC LITERACY AND CAUTION

A human's ability to use maps is to some degree a learned behavior. Humans learn universal cartographic symbols and conventions that make understanding an unfamiliar map easier (Allan, 1999). For instance, some small children have difficulty with the concept that a truncated map boarder does not display the literal shape of countries that extend past the map's edge (Arnheim, 1969). In other words, students must learn that the items or concepts mapped have the potential to extend past the edges of the representational image.

Literacy with KDV's must also be cultivated. A user must first become familiar with the metaphor that spatial proximity, while seldom literal, implies proximity between the concepts being mapped. Once understood, the use of the next domain map should be easier. Also, the establishment of universal symbols and conventions should likewise increase the speed in which a domain map is understood. While efforts may be made to enunciate best practices and to establish conventions, these items will most likely emerge over time through the repeated creation and use of domain maps and user testing.[3]

There are also risks in using KDVs in education. KDVs may convey to the viewer a false sense of established proximities when in fact domains are fluid. In addition, there may be different paradigms or viewpoints for visualizing a domain. Also, different techniques used to produce KDVs will create different representations of a knowledge domain. In other words, there are no deterministic, exactly reproducible methods for creating KDVs. "Even with clear assumptions and good qualitative research methodology, there are a myriad of ways to create a single type of map of the same content. It is important to either involve a group of scholars in developing a map and/or to recognize that the resulting map is simply one of many possible patterns for documenting the links between complex ideas" (Milam, Santo, and Heaton, 2000, p. 5). While Milam is referring to concept maps created without the more rigorous automated techniques employed by information scientists, the warning is still applicable. An educator will have to consider the risks and benefits of using a single KDV or presenting multiple KDVs of a particular domain created with different techniques or that encapsulate different paradigms.

[3] Ware cautions: "standardization is the enemy of innovation and innovation is the enemy of standardization" (Ware, 2004 p. 386).

13. FUTURE DIRECTIONS FOR EDUCATIONAL KDVs

The authors envision six phases in the use, adoption, and implementation of KDVs. The first two have already occurred. The remaining four are prospective—compelled by the cognitive benefits discussed above.

Phase one: bibliometricians realized that they could use bibliographic datasets and techniques such as author co-occurrence to provide maps of a particular discipline. These early maps were graphically simple and painstakingly created by hand.

Phase two has been the implementation of automated techniques for data harvesting, processing, and information visualization. This has facilitated the mapping of larger domains.

Phase three: domain maps will become widely known outside of information science (Whitehouse, 2004). They will become popular with educators and will be used to enhance classroom pedagogy. In April 2005, a workshop at the annual conference at the American Association of Geographers brought together cartographers and information scientists to discuss creating better KDVs. See http://vw.indiana.edu/aag05/.

Phase four: the widespread use of domain maps will lead to steps that aim to harmonize and better preserve the scholarly data from which they are created (Börner, in press). This might include unique author identification numbers, better and standardized citation practices, and repositories containing information such as all of the sources cited in books.

Phase five: domain maps will routinely be used as one of the access options to digital libraries and online public access systems (OPAC's) (the tools library patrons use to search for materials). Even if a user chooses to do a keyword search, domain maps will be dynamically displayed unobtrusively in the background, subtly conveying to the user the intellectual landscape of the particular domain and the specific neighborhood of the user's search interest.

Phase six: dynamic domain maps will capture and portray the diffusion of information. This diffusion of knowledge may occur from one author to the next, one journal to another, or among scientific disciplines, etc. Domain maps will have predictive elements that will forecast and model the spread of knowledge. They will be used widely for science forecasts in a similar fashion as today's weather forecast maps.

14. CONCLUSION

The colossal landscape of scholarly knowledge, growing at exponential rates, now requires representational maps utilizing advanced techniques to provide insight into the structure and dynamics of scholarly domains. Today we need intellectual cartographers to assist students and scholars in navigating, understanding and internalizing the structure and dynamics of scholarly bodies of knowledge. There is compelling evidence of the utility of KDVs stemming from the fields of educational psychology, cognitive science, cartography, and information science. Well designed KDVs have the ability to facilitate understanding, recall, and to convey to the user the schematic, geo-spatial, temporal, semantic, or social organization of the underlying domain.

Though educational knowledge domain visualizations are still in their infancy, we believe that they have a promising future in assisting with access to and the navigation, understanding, management, and communication of large-scale information spaces. Furthermore, when used as an interface for information retrieval, knowledge domain visualizations have the potential to convey the structural organization of the domain to the user. In turn, this structural knowledge of the domain provides the cognitive scaffolding with which the user may associate additional details about the domain.

REFERENCES

Allen, G. L. (1999). Spatial Abilities, Cognitive Maps, and Wayfinding: Bases for Individual Differences in Spatial Cognition and Behavior. In: R. G. Golledge (Ed.), *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes,* pp. 46–80. Baltimore, MD: Johns Hopkins University Press.

Arnheim, R. (1969). *Visual Thinking.* Berkeley, CA: University of California Press.

Bernal, J. D. (1939). *The Social Function of Science.* London: Routledge and Kegan Paul Ltd.

Bertin, J. (1983). *Semiology of Graphics: Diagrams Networks Maps.* Translated from the French by William J. Berg. Madison, Wisconsin: University of Wisconsin Press.

Börner, K. (in press). Semantic Association Networks: Using Semantic Web Technology to Improve Scholarly Knowledge and Expertise Management. In: V. Geroimenko and C. Chen (Eds.), *Visualizing the Semantic Web*, Springer Verlag, 2nd Edition, Chapter 11.

Börner, K., Chen, C., and Boyack, K. W. (2003). Visualizing Knowledge Domains. In: B. Cronin (Ed.), *Annual Review of Information Science and Technology* (pp. 179–255). Medford, NJ: Information Today.

Braam, R. R., Moed, H. F., and Van Raan, A. F. J. (1991a) (1). Mapping of Science By Combined Co-Citation and Word Analysis. I. Structural Aspects. *Journal of the American Society for Information Science,* 42(4) 233–251.

Braam, R. R., Moed, H. F., and Van Raan, A. F. J. (1991b) (2). Mapping of Science By Combined Co-Citation and Word Analysis. II. Dynamical Aspects. *Journal of the American Society for Information Science,* 42(4), 252–266.

Breuker, J. A. (1984). A Theoretical Framework for Spatial Learning Strategies. In: C. D. Holley and D. F. Dansereau (Eds.) (1984). *Spatial Learning Strategies: Techniques, Applications, and Related Issues.* Orlando, FL: Academic Press.

Buter, R. K., Noyons, E. C. M., and Van Raan A. F. J. (2004). A Combination of Quantitative and Qualitative Maps in an Evaluative Bibliometric Context. In *Proceedings of the Eighth International Conference on Information Visualisation (IV'04)*, (pp. 1093–1095) 47/04.

Callon, M., Law, J., and Rip, A. (Eds.) (1986). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World.* London: Maxmillan Press Ltd.

Chen, C. (2003). *Mapping Scientific Frontiers: The Quest for Knowledge Visualization.* New York: Springer.

De Solla Price, D. J. (1965). Networks of Scientific Papers *Science*, New Series, 149(3683) (Jul. 30, 1965), 510–515.

Feather, R. M. Jr. Snyder, S. L. and Hesser, D. T. (1993). *Concept Mapping, Workbook to Accompany, Merrill Earth Science*. Lake Forest, Illinois: Glencoe.

Garfield, E. Sher, I. H. Torpie, R. and Torpie, J. (1964). *The Use of Citation Data in Writing the History of Science.* Philadelphia: Institute for Scientific Information.

Garland, K. (1994). *Mr. Beck's Underground Map*. London: Capital Transport Publishing.

Gastner, M., Shalizi, C., and Newman, M. (2004). Maps and Cartograms of the 2004 US Presidential Election Results. Available at: http://www.cscs.umich.edu/~crshalizi/election/

Golledge, R. G. (1999). Human Wayfinding and Cognitive Maps. In: G. R. Golledge (Ed.), *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes,* pp. 5–45. Baltimore, MD: Johns Hopkins University Press.

Holley, C. D. and Dansereau, D. F. (Eds.) (1984). *Spatial Learning Strategies: Techniques, Applications, and Related Issues.* Orlando, FL: Academic Press.

Hook, P. A. (2002). Creating an Online Tutorial and Pathfinder. *Law Library Journal,* 92(4), 243–265.

ISI (Institute for Scientific Information) (1981). *ISI Atlas of Science: Biochemistry and Molecular Biology, 1978/80*, Philadelphia, PA.: Institute for Scientific Information.

Jonassen, D. H., Veissner, K., and Yacci, M. (1993). *Structural Knowledge: Techniques for Representing Conveying, and Acquiring Structural Knowledge.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kealy, W. A. (2001). Knowledge Maps and Their Use in Computer-Based Collaborative Learning Environments. *Journal of Educational Computing Research,* 25(4) 325–349.

Koehler, W. (1947). *Gestalt Psychology* (Rev. Ed.). NY: Liveright.

Kruskal, J. B. and Wish, M. (1984). *Multidimensional Scaling*. London: Sage Publications.

Kulhavy, R. W. and Stock, W. A. (1994). Reference Maps as a Framework for Remembering Text. In: W. Schnotz and R. W. Kulhavy (Eds.), *Comprehension of Graphics* (pp. 153–162). Amsterdam: North-Holland.

MacEachren, A. M. (1995). *How Maps Work: Representation, Visualization, and Design.* New York: Guilford Press.

MacEachren, A. M., Gahegan, M., and Pike, W. (2004). Visualization for Constructing and Sharing Geo-Scientific Concepts. *PNAS*, 101(Suppl_1): 5279–5286.

Mane, K. and Börner, K. (2004). Mapping Topics and Topic Bursts in PNAS. *PNAS,* 101(Suppl_1): 5287–5290.

Marshakova, I. V. (1973). A System of Document Connections Based on References. *Scientific and Technical Information Serial of VINITI,* 6, 3–8.

McCain, K. W. (1990). Mapping Authors in Intellectual Space: A Technical Overview. *Journal of the American Society for Information Science,* 41(6) 433–443.

Milam, J. H., Jr., Santo, S. A., and Heaton, L. A. (2000). *Concept Maps for Web- Based Applications.* ERIC Technical Report. ERIC No.: ED445627.

Miller, G. A. (1968). Psychology and Information. *American Documentation.* July, 286–289.

Montello, D. R., Fabrikant, S. I., Ruocco, M., and Middleton, R. S. (2003). Testing the First Law of Cognitive Geography on Point-Display Spatializations. In: W. Kuhn, M. F. Worboys, and S. Timpf (Eds.), *Spatial Information Theory: Foundations of Geographic Information Science.* In *Proceedings of the Conference on Spatial Information Theory (COSIT '03). Lecture Notes in Computer Science, 2825* (pp. 316–331). Berlin: Springer Verlag.

Montello, D. R., Hegarty, M., Richardson, A. E., and Waller, D. (2004). Spatial Memory of Real Environments, Virtual Environments, and Maps. In: G. L. Allen (Ed.) *Human Spatial Memory: Remembering Where.* Mahwah, New Jersey: Lawrence Erlbaum Associates, 251–285.

Moore, P. and Fitz, C. (1993). Gestalt Theory and Instructional Design. *Journal of Technical Writing and Communication,* 23(2), 137–157.

Nesbitt, K. V. (2004). Getting to More Abstract Places using the Metro Map Metaphor. In *Proceedings of the Eighth International Conference on Information Visualisation (IV'04).* Piscataway, NJ: IEEE.

Newbern, D., Dansereau, D. F., and Patterson, M. E. (1997). Spatial-Semantic Display Processing: The Role of Spatial Structure on Recall. *Contemporary Educational Psychology,* 22, 319–337.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Rubin, R. E. (1998). *Foundations of Library and Information Science.* New York: Neal-Schuman Publishers, Inc.

Skupin, A. (2004). The World of Geography: Visualizing a Knowledge Domain With Cartographic Means. *PNAS*, 101(Suppl_1): 5274–5278.

Skupin A. and Fabrikant, S. I. (2003). Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. *Cartography and Geographic Information Science*, 30(2), 95–119.

Slocum, T. A. (1998). *Thematic Cartography and Visualization*. Upper Saddle River, New Jersey: Prentice Hall.

Small, H. (1973). Co-Citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. *Journal of the American Society for Information Science,* 24, 265–269.

Small, H. (1999). Visualizing Science by Citation Mapping. *Journal of the American Society for Information Science,* 50(9), 799–813.

Thorndyke, P. W. (1981). Distance Estimation From Cognitive Maps. *Cognitive Psychology,* 13, 526–550.

Thorndyke, P. W. and Hayes-Roth, B. (1982). Differences in Spatial Knowledge Acquired From Maps and Navigation. *Cognitive Psychology,* 14, 560–589.

Ware, C. (2004). *Information Visualization: Perception for Design.* San Francisco, CA: Morgan Kaufman.

West, C. K., Farmer, J. A., and Wolff, P. M. (1991). *Instructional Design: Implications from Cognitive Science*. Englewood Cliffs, New Jersey: Prentice Hall.

White, H. and McCain, K. (1998). Visualizing a Discipline: An Author Co-citation Analysis of Information Science, 1972–1995, *Journal of the American Society for Information Science*, 49(4), 327–356.

Whitehouse, D. (2004). *Scientists Seek 'Map of Science.'* BBC News. Wednesday, 7 April, 2004. Available at: http://news.bbc.co.uk/1/hi/sci/tech/3608385.stm.

Winn, W. and Holiday, W. (1982). Design Principles for Diagrams and Charts. In: D. H. Jonassen (Ed.). *The Technology of Text: Principles for Structuring, Designing, and Displaying Text* (pp. 277–299). Englewood Cliffs, NJ: Educational Technology Publications.

Winn, W. (1994). Contributions of Perceptual and Cognitive Processes to the Comprehension of Graphics. In: W. Schnotz and R. W. Kulhavy (Eds.), *Comprehension of Graphics* (pp. 3–27). Amsterdam: North-Holland.

CHAPTER 11

LEARNING AND TRAINING TO SEARCH

WENDY LUCAS AND HEIKKI TOPI
Bentley College

1. INTRODUCTION

Online searching was once the province of professional search intermediaries who had received formal training in information retrieval (IR). Today's searchers no longer fit a standard profile, ranging from school-age children to retirees, the majority of whom have received no training in search. While entering a simple query to a commercial Web search engine may not require any specialized skills, a basic understanding of the search process and successful search techniques and strategies, particularly for non-trivial queries, can result in more productive searches and a more enjoyable search experience. Despite recognition of the growing need for search training that supports the cognitive processes and learning styles of individual searchers, there has been relatively little response from the academic community to date.

In this chapter, we review the existing literature on search training. We also propose the development of a search training framework based on models of the search process, research on search errors, and models, methods and approaches gathered from research on training and learning within other, related disciplines.

2. BACKGROUND

The main focus of this chapter is on training that is intended to help those whose main profession is not performing information searches to improve their online search skills and their understanding of the search process. In our context, online search includes a number of different search environments: the public Web, proprietary databases available either through the Web or on private networks, online public access catalogs, corporate knowledge management systems, etc. The ability to use online search tools effectively has become an integral part of every knowledge professional's skill set and, therefore, it is critically important to understand how these skills can be best developed.

Recent research from Pew (Fallows, 2005; Fallows, Rainie, and Mudd, 2004) indicates that 84 percent of Americans who have used the Internet have also used one or more search engines. According to this study, more than 38 million Americans use a search engine every day, and over 90 percent feel either very or somewhat confident about their abilities as users. As will be discussed later in this section, academic research suggests that this confidence is not always well-founded. This is also supported by additional findings by Pew (Fallows, 2005): for example, 62 percent of the searchers were not aware of the general difference between paid and unpaid search results, which raises questions about their general understanding of the search environments being used.

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 209–226.

The information-seeking process has most notably been characterized by Kuhlthau's Search Process Model (Kuhlthau, 1993), which defines six stages of information seeking: task initiation, topic selection, prefocus exploration, focus formulation, information collection, and search closure. These findings are based on studies of library patrons and do not include IR system usage. Marchionini (1992) divides the information-seeking task of IR system end users into five functions: defining the problem, selecting the source, articulating the problem, examining the results, and extracting the relevant information. While this is more closely aligned with our interests, our focus is still more narrowly defined on information searching, rather than the broader context of information seeking (Bates, 1985).

The end user, in this case, is the searcher who is already cognizant of the information need and available search repositories and is faced with the task of searching via queries for answers to that need. Thus, the process of finding search repositories and choosing between them lies outside of our focus. We are interested in identifying the skills searchers can learn and methodologies for teaching those skills in order to increase their effectiveness at searching.

Studies have shown that a searcher's ability to get desired results is often very weak (Spink et al., 2001; Lucas and Topi, 2002). Key among these findings is the inability of searchers to choose correct search terms and to correctly use operators. It is well known that searches tend to include very few terms in their queries and seldom include Boolean operators, and that the correct usage of those operators is even less frequent. Taking a higher level perspective, searchers are often not aware of a good strategy to pursue for a successful search process and have little knowledge of the intricacies of that process.

At the same time, search is increasingly important for both individuals and organizations because of the growing number of electronically stored non-structured and semi-structured documents, such as corporate memos, PowerPoint™ presentations, Web blogs, online news content, knowledge management resources, etc. While the amount of available information has continued to grow exponentially, advances in human ability to store and process information have not been equally impressive. Therefore, online search environments have become essential tools, and it is vitally important for a knowledge professional to be able to use them effectively.

3. MOTIVATION

The ubiquitous availability of the Web and other online data sources has created a situation in which, in practice, all knowledge professionals have to be capable searchers in order to perform their primary job functions well. The designator of "knowledge professional," in this context, covers a large swath of the population, such as:

- A marketing executive looking for information about competitor's products.
- A development director at a non-profit searching answers to questions about a new target country.
- A low-income citizen using the Web to search for information about earned income tax credit.
- A middle-school student writing a research paper.
- A sports enthusiast investigating a career of a favorite player.

Clearly, the need to perform effective searches is not limited to those using these tools in professional contexts; there are several personal reasons to become an effective user of search tools as well.

Increasingly powerful search engines can solve only part of the problem when it comes to search effectiveness; helping individual searchers develop their search skills is another essential part of the equation and is the primary motivation for this chapter. A variety of different systems with varying capabilities are available, depending on the application. While some, such as most commercial Web search engines, may appear to be easy to use effectively with little experience or knowledge about searching required, others, such as those found on many corporate intranets and those used by digital libraries, may require considerable understanding of the search process, including query formation and use of system features, in order to be used productively. Research has shown that:

- Searchers often don't realize that their search skills are relatively weak (an illustrative anecdote is that, in a survey of about 150 freshmen in a business undergraduate program taking an introduction to technology course, the self-reported search skills level was 4.7/5, while an experimental study using the same population a few years earlier indicated serious difficulties in searching, particularly with complex queries).
- Searchers tend to have very little time and patience for search training, largely because they believe that they are already good at searching.
- Web search engines' promise of producing the best possible answer set on the basis of very simple queries contributes to the discrepancy between searchers' confidence and their true abilities.
- Searchers tend to avoid search interfaces that go beyond a simple text box, partially because the more complex interfaces are often labeled as "advanced," even though they would be very helpful for inexperienced users.

It is not always clear whose responsibility it should be to offer training in search. The ability to do successful online searches has become a modern literacy skill, and all universities and colleges should, therefore, do their best to make sure that the search skills of their graduates are as well developed as possible. Businesses, governments, and non-profit organizations also benefit from having employees who are highly capable searchers.

As will be discussed in the next section, there is very little existing academic literature on search training, and it appears that there have been few systematic research programs focusing on this topic (with the exception of the University of Tampere program reported, for example, in Halttunen (2004)). This is despite Borgman's (1996) call for improved training and documentation based on an in-depth understanding of information seeking behavior. A challenge therefore exists for those who would like to get guidance from the academic community on understanding how to build better search support programs.

It is, however, possible that we might be able to find such guidance from other academic disciplines that have dealt with similar types of issues. Prior research that has focused on the factors underlying successful search will also be helpful, because it will allow us to identify the key areas for both practical search training programs and future research on search learning and training.

4. REVIEW OF EMPIRICAL RESEARCH

We will continue this chapter by evaluating the existing literature on online search training. As mentioned above, the amount of existing research is relatively modest, and it appears to be based on discrete projects instead of systematic research programs. There is one notable exception: Halttunen (2004) and the studies that are included in this dissertation (Halttunen, 2003a, b; Halttunen and Järvelin, 2004; Halttunen and Sormunen, 2000). We will first describe the relevant studies and then review the conclusions we can draw based on this research.

Detlor (1999), following Davenport and Prusak (1993), suggests that corporate libraries should break away from their traditional role as information custodians and move towards an information broker role, in which their main task is to help build connections between those who have information (humans or databases) and those who do not. He makes the point that part of this process is providing effective Web search training. This case study of one company provides evidence supporting the idea that search training is capable of helping searchers improve their search skills, particularly by providing, on the one hand, very detailed instructions (such as how to use shortcuts) and, on the other hand, broad search strategies.

The participants in this training program also indicated that they benefited from exposure to tools of which they had previously been unaware, such as meta search engines, and from a review of available internal resources, such as a corporate Intranet site. The participants found hands-on training that was closely linked to their work practices to be particularly helpful. Three major benefits were identified: more effective information searches, increased participant confidence, and increased appreciation of the organizational unit that provided the training. Unfortunately, the study does not describe at a detailed level the nature of the training that was provided and, therefore, it is difficult to determine what contributed to the success of search in this context.

Cloyd and Spilker (2000) investigated the effects of the searcher's academic education on confirmation bias in the search process. Confirmation bias in the context of tax research refers to the extent to which the person conducting the research (and, consequently, the searches) focuses on information that supports a position favorable to the client and ignores negative findings. Cloyd and Spilker (2000) found that law students are less prone to this bias than accounting students. Even though this finding is not specifically related to a short-term training process, it is interesting because it suggests that the type of domain expertise a searcher has gained by training may have a profound impact on the approach to searching.

Lazonder (2000) studied the training needs of a small number of teenage Web searchers through a detailed analysis of the differences between novice and experienced searchers. He specifically identifies the difference between these two user types as a "classical yet fruitful" way to analyze the training needs of the novice group. This study did not find highly significant differences between the groups, particularly in the later stages of a multi-task experiment. Lazonder's (2000) suggestion is that training should focus on specific search engine features that are often not known to novices and on Boolean operators. Also, training should help the users to use feedback data provided by the search engine (such as relevance ratings and additional keywords suggested by the system).

Debowski (2002) also acknowledges the lack of research on search and information seeking (a broader process that provides the context for search) training. She suggests that an information seeking process consists of a series of potentially highly complex decision processes and recognizes that decisions regarding information search have a strong impact on the eventual domain-specific decisions. Also, she categorizes problems in search into three groups: a) selection of keywords (search terms), b) use of command structures (operators), and c) creation of a cohesive, integrated search process.

In addition, Debowski (2002) reports the results of two search training studies. The first study found that a training program based on the ideas of "structured guided mastery" (Bandura, 1986) was, in this search context, significantly more effective than a program that was based on unguided, enactive exploration. The author suggests that this result might have occurred because of the complexity of the search context. The results of the second study indicate that enactive exploration training provided after guided mastery training has a positive impact on intrinsic motivation but no effect on overall quality of the search.

Debowski (2002) also describes an experiment that evaluated the effects of process-related feedback on performance in an information search task. The findings strongly support the value of providing a novice searcher with both task and search strategy validity feedback. All of the experiments reported in Debowski's (2002) study testify about the need to develop effective search training approaches; even after training, the performance level in these experiments was quite low, as it often is with complex search tasks (see, for example, Sutcliffe, Ennis, and Watkinson (2000) and Lucas and Topi (2004)). Debowski (2002) also points out the importance of integrating training and ongoing support.

Cheng (2003) studied the effectiveness of a three-hour workshop as a mechanism for improving the search skills of medical professionals. The workshops were divided into four parts, each consisting of conceptual material, live demonstrations, and hands-on practice. During the course of the workshop, the participants covered eight exercises, each of which required them to address a clinical search scenario. He evaluated the success of the workshops with four different dependent variables, following Kirkpatrick's (1998) model: immediate satisfaction, learning, behavior, and results. This carefully designed study with a large sample suggests that the workshop was effective in improving the participants' skills in question formulation, raising confidence in their ability to search, developing search skills, and improving search outcomes.

Lucas and Topi (2004) and Topi and Lucas (2005) evaluated the effectiveness of brief computer-based training interventions on the searcher's ability to retrieve correct answers from a simulated intranet site with a commonly used search engine. Their results suggest that even a modest Boolean training intervention can be very effective in cases where the user interface does not provide the support that the users' expertise level would warrant.

As mentioned earlier, by far the most comprehensive and systematic effort to evaluate search learning and training known to us is a multi-study project by Halttunen (2004). This dissertation reports the design and evaluation of two IR learning environments intended to support an academic course. It also provides a thorough analysis of IR literature, to the extent that it is relevant for the main focus of the study, and it discusses the applications of various learning models to this context. We will briefly review the main findings of the component articles of this thesis.

Halttunen and Sormunen (2000) describe and evaluate an early version of the IR Game, a pedagogical game intended for IR education. This article serves mostly as a precursor and a developmental step towards the more comprehensive main study reported in the other three articles. In a later article, Halttunen (2003a) focuses on the participants' perceptions regarding the IR process; the purpose of this analysis was to provide background material for the development and use of the IR learning environments. The article emphasizes the differences between individual perceptions regarding the IR process and the effects of these differences on IR learning and teaching.

Further, Halttunen (2003b) evaluates the effectiveness of two pedagogical approaches, namely scaffolding and anchored instruction, in an introductory IR course. Scaffolding refers to (gradually fading) supporting interventions by persons with higher levels of expertise and/or experience, which allow a learner to perform tasks that they could not complete without that support. Anchoring requires that the instructional approach be situated in a meaningful problem solving context. Several noteworthy findings emerge from this study: 1) the importance of using relevant, meaningful learning tasks that provide a context for the learning process; 2) the negative effect of decontextualized tasks on motivation; 3) the importance of closely linking theory and practice; and 4) the importance of feedback.

Halttunen and Järvelin (2004) report the results of the learning outcome comparison between a traditional group and an experimental group that had the benefit of systematically implemented scaffolding and anchoring. The experimental group performed fewer semantic knowledge errors (errors in transforming an information request into a query) and were able to achieve better search results.

4.1. Key Findings

Many of the articles reviewed above present a statement on the status of the research on search/IR learning and/or training. The tone of all of them is similar, as can been seen from these quotations:

> "Research on the instruction of IR is disjointed, lacking of a solid background in both information studies and education research" (Halttunen, 2003a, p. 324).
>
> "Yet, there is little research on effective instructional methods for teaching users how to search the Web" (Colaric, 2003, p. 111).
>
> "Despite the growing importance of these information sources, little is known about the process of information seeking, and the ways in which skills may be improved" (Debowski, 2002, p. 21).

The consensus in the literature is that research on search learning and training is still in its very early stages of development. Collectively, the studies reviewed here demonstrate that, particularly with non-trivial information requests, successful searching requires a combination of complex cognitive skills that have to be learned. This learning process can be supported with suitable pedagogical interventions. Interesting, consistent (or at least not contradictory) findings include the following:

- In general, the search skills of the users are weak and can be improved by even brief training interventions, as long as those interventions are effectively targeted.

- Many searchers benefit from relatively small details that help them use the technical search environment more effectively or access data sources that they were not aware of earlier.
- The effectiveness of a training approach is dependent on a number of factors, such as complexity of the task, the type of search technology being used, the level of the user's expertise, and other individual user characteristics.
- Participants find the most effective training approaches to be those that integrate hands-on training and conceptual material. In addition, it is beneficial if the training is implemented using material from a domain that is meaningful to the trainees. It appears that, when tested in IR, methods and approaches that have been found useful in general educational literature (such as scaffolding and anchoring in Halttunen's work) are effective.
- Training should help the users understand and utilize the feedback that they receive from the search engine.

At present, there are no well-defined, cohesive theoretical findings emerging from the existing search training literature. Halttunen's dissertation (2004) is currently the best effort to integrate relevant literature from the reference disciplines (particularly education) with an in-depth IR understanding, and this work should be commended and more widely distributed.

5. TRAINING FOR SEARCH

In this section, we discuss three possible ways to provide structure to future efforts to study the factors affecting search training. First, we evaluate the role search process models could play in helping us better understand the requirements for search training. Here, the core assumption is that search training requirements can best be understood by utilizing the in-depth knowledge that research on search processes has produced. Second, we choose a somewhat different perspective and review research on search errors. One of the key goals of training should be helping the trainees to avoid errors. To achieve this goal, it is necessary to understand the characteristics of search errors and the reasons underlying them. Finally, we briefly explore how researchers in a related field, computer software training, have modeled training and evaluate whether or not some of the research lessons learned there can be applied to search training.

5.1. Search Process Models and Search Training

An understanding of the search process is essential for understanding search training requirements because, in practice, the goal of training should be to improve performance either in any one of the individual process stages or in the user's ability to perform an integrated combination of those stages. If it is not clear what capabilities the training is targeted to improve, it is very difficult to determine whether or not it has been successful.

Borgman (1996, p. 495) has identified three layers of knowledge required for online catalog searching that can serve as a framework for the search learning process:

- Conceptual knowledge of the IR process—translating an information need into a searchable query;

- Semantic knowledge of how to implement a query in a given system—the how and when to use system features;
- Technical skills in executing the query—basic computing skills and the syntax of entering queries as specific search statements.

Stage 1:	Articulate information need.
Stage 2:	Conceptualize need as an executable query.
Stage 3:	Formulate the conceptual query for the given system.
Stage 4:	Enter the query to the search interface.
Stage 5:	Understand and interpret the results.

Figure 1: Search process stages

These layers can also be mapped to the stages of the search process, which are defined in Figure 1. We should emphasize again that we are focusing on a relatively narrowly defined search process in which a person has an information need that he desires to fulfill instead of the broader information seeking process. It is also important to note that searching is conducted as an iterative process in which a searcher may make refinements to any prior step as deemed necessary within a search session. A session is defined here as the entire sequence of queries entered by a searcher for each information request (Jansen and Pooch, 2001). Feedback from previous queries within a session may and often do have an impact on the formation of the later queries.

Our model bears similarities to Marchionini's (1992) five stage model, namely: defining the problem, selecting the source, articulating the problem, examining the results, and extracting the relevant information. While this model is somewhat aligned with our interests, our focus is more narrowly defined on information searching. The first two stages in Marchionini's model have been completed prior to the first stage in our model. Our model specifically addresses the gap between his stages of problem articulation and examination of results.

During the first stage, the searcher must conceptualize the information that is actually being sought. This is a critical step, for if the information need is not correctly articulated, the final outcome of the following stages cannot possibly be successful. Next, the searcher must conceptualize that articulated need as an executable query. This involves the ability to identify relevant search terms, which is actually a more complex process than the typical searcher may realize. Relevancy is evaluated by the search engine in terms of the content of documents in the repository being searched that are likely to meet the information need, as expressed by the query, rather than in terms of the actual question to which that searcher is seeking an answer.

Furthermore, the searcher must understand how those terms can be combined, which requires some conception of Boolean logic. The process of translating a user request into a Boolean query has long been recognized as a particularly difficult one for untrained searchers (Cooper, 1988; Ingwersen and Willet, 1995), with general confusion reigning over the logical difference between the AND and OR operators (Borgman, 1996). These first two stages pose significant difficulties for novice searchers, despite the fact that they can take place offline because they are independent of the search system being used.

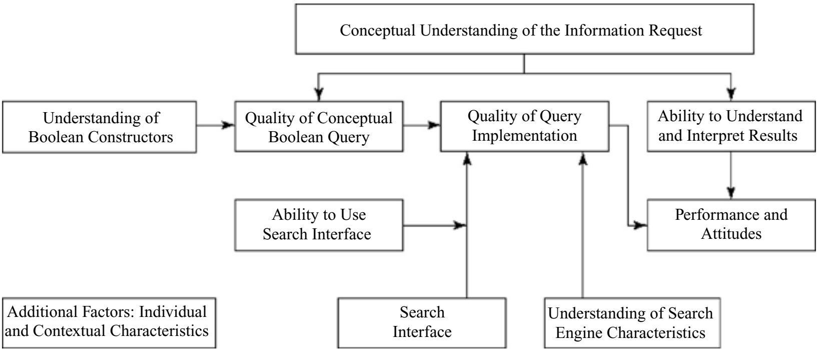

Figure 2: Learning domain model

In the third stage, the searcher must formulate the query for the given search system. To fully take advantage of the system's potential requires knowledge of available search features, i.e., which operators are supported, operator order of precedence, interpretation of search term ordering, etc. These implementation specifics must be searched for by the system user, something that inexperienced searchers are unlikely to do given that most have no knowledge of even the existence of such options, let alone their application.

The searcher must then be able to correctly enter the query information to the search interface. This entails having general computer usage knowledge as well as familiarity with interface components, such as text boxes, buttons, drop-down lists, etc. After taking action to execute the query, the searcher must then know how to interpret the results, which is based on system-specific information, such as ordering of results by relevancy; computer usage familiarity, such as using a find feature to search within documents; human intelligence, and other personal characteristics.

Figure 2, which first appeared in (Lucas and Topi, 2004), models the learning domain by those human-dependent factors that ultimately affect the success of a search. Search success is defined here, as per Chan, Tan, and Wei (1999) and Te'eni and Feldman (2001), in terms of search performance (correctness/accuracy, time) and searchers' attitudes (confidence and satisfaction). This model identifies the areas in which searchers can learn how to improve their performance, in accordance with the knowledge requirements of the search process stages identified in Figure 1.

Also included in this model is the search interface itself, which is intended to assist the searcher in implementing a query but may actually hinder that process by being a source of confusion (Shneiderman, Byrd, and Croft, 1997). Additional factors are the numerous individual and contextual characteristics that have been shown to affect all stages of the search process. These include, but are not limited to, the searcher's experience and domain knowledge (Hölscher and Strube, 2000), cognitive abilities and style (Allen, 2001; Ford, Miller, and Moss, 2001), problem-solving style (Kim and Allen, 2002), self-efficacy (Ford, Miller, and Moss, 2001), maturity, and ability to concentrate. Failure to take these factors into account when training users on search skills can prove fatal to the success of that training.

With these two models as an example, we can clearly see the benefits of using a search process model as a foundation for the development of a search training model. Steps in a search process model can easily be translated into areas in which training is needed. Obviously, it must also be acknowledged that both interaction between individual stages and the entire search process need coverage in training, but a detailed understanding of the search process undoubtedly provides support in identifying training needs.

The model in Figure 1, for example, directly reveals the requirement to train users in articulating their information needs, translating information needs into queries (first conceptually and then in an executable form), physically running the query, and processing the results. The model itself does not directly tell how training in these areas should be performed, but the better we understand a specific process step, the easier it becomes to design a training intervention. A process analysis approach to understanding the training needs also reveals the major differences between the training requirements for the different stages: training in the articulation of information needs and training in the utilization of the search results are closely linked to improving the searcher's conceptualization of the domain within which the search takes place, whereas training related to the execution of a query within a specific environment is very similar to other types of technical computer training. Translation of information needs into queries is fundamentally a translation process between modeling domains at different levels of abstraction and, as such, is likely to require a different training approach.

The second model (Figure 2) leads to the identification of a somewhat different, but not less relevant, set of training needs. It clarifies and further specifies the stages introduced in the first model. For example, it identifies a specific need to train users in the use of Boolean constructors (part of the information need → query translation stage), the characteristics of the search interface (execution stage), and the idiosyncrasies of the search engine being used (information need → query translation and execution stages).

As discussed in Pharo (2002), a large number of other models have been developed to describe various aspects of the information seeking or searching process or behavior, including those by Marchionini (1992), Kulthau (1993), Ingwersen (1992, 1996), and Spink and Saracevic (1997). In this research, we chose to use our own models as a basis for demonstrating the value of a search framework for modeling training needs because of the straightforward process nature of our model. For example, Ingwersen's (1996) comprehensive cognitive interactive IR model does not specify process stages, and Kulthau's (1993) model is at a higher level of abstraction. As previously discussed, our five stage model focuses at a more detailed level on the search process itself than does that of Marchionini (1992).

5.2. Lessons from Research on Search Errors

The prior section points to the need for search skill training, as opposed to general computer usage training, in the areas of term identification, Boolean operator usage, search interface usage, and query refinement. The last is based on interpretation of output, including feedback provided by the search engine in the form of relevancy ratings and keyword suggestions (Lazonder, 2000).

In identifying the critical areas for search training, it is useful to investigate the behaviors of searchers and the types of errors most commonly made during the search process. These errors can then be viewed from the perspective of Borgman's (1996) knowledge framework and analyzed in terms of their effects on the stages of the search process, as defined in Figure 1.

Several studies have evaluated queries logs of commercial Web search engines (see, for example, Abdulla, Liu, and Fox, 1998; Jansen, Spink, and Saracevic, 2000; Silverstein et al., 1999) and reached similar conclusions concerning the contents of queries. Spink et al., (2001) analyzed over one million queries submitted by over 200,000 users to the Excite search engine and found that the average number of terms per query was two and that fewer than 5 percent of all queries used any Boolean operators, with the AND operator being the most commonly used (in 3 percent of queries), followed by the OR operator (in 1 percent).

The most frequent mistake made with Boolean operators was not capitalizing them, as required by Excite. While the use of the plus ("+") and minus ("−") modifiers (in 7 percent of all queries) exceeded that of the Boolean operators, most usage was incorrect (in 5 percent of all queries), as determined by conformance with Excite's search rules. Another category of error was that of using searching techniques that were not supported, such as including SEARCH as an operator, the semi-colon (";") as a separator, the period (".") as a separator or as part of a URL or e-mail address, and the ampersand ("&") to represent an AND condition.

These findings confirm the need for training on operator usage, both independently and in conjunction with the search tool being used. They also point to a lack of conceptual and semantic knowledge, which will affect the query conceptualization and formulation stages identified in Figure 1. While log studies provide great insight into the common characteristics of Web queries, including typical errors, their findings cannot be related to their effect on performance, as the information needs of the searchers and the successfulness of their searches cannot be determined.

A study by Sormunen and Pennanen (2004) identifies common errors based on 1,037 queries to a traditional Boolean IR system (TRIP) and 388 to a best-match system (InQuery). In addition, it relates those errors to the performance of the queries in which they occurred. Two types of analysis were completed, one of low-performance queries, which were defined as those not exceeding a level of 10 percent in average uninterpolated precision, and the other for complete search sessions by two user groups at each performance extreme (i.e., one with high success rates and the other with low rates). Errors were evaluated in terms of Borgman's categorization.

Figure 3 shows the types of error associated with each of these categories.

The low-performance queries helped identify errors that have a serious effect on search effectiveness. For these queries, users failed more frequently in Boolean than in best-match queries. About 70 percent of errors in Boolean queries related to conceptual knowledge, with the majority (37 percent) from *weak facet applied*. About 13% of errors were attributable to semantic knowledge, and the remaining 16% were technical errors, of which 14 percent were attributable to *common syntax errors*. For the best-match queries, most of the errors were technical (85 percent). Technical errors are usually associated with Boolean search systems, which are exact-match, rather than best-match systems, but the data in this particular analysis was biased due to what the authors refer

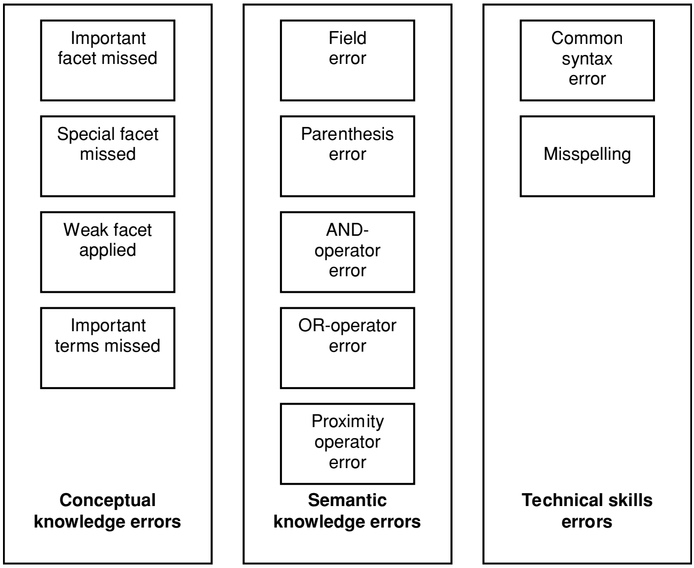

Figure 3: Error categorization model (Based on Sormunen and Pennanen, 2004)

to as an immaterial feature of the database index of which the searchers were not aware.

The analysis of search sessions found similar results for the Boolean queries in terms of error distribution. For the best-match queries, the majority of errors were in the conceptual knowledge category, though those errors did not have a substantial effect on performance. Technical and semantic errors were found to be the best single explanation for low performance in these types of queries for both types of analysis performed in this study.

It is interesting that the main causes of low search performance in the Boolean system were conceptualization errors, which correspond to Stage 2 of the search process, as modeled in Figure 1. As noted by Lazonder (2000), while sophisticated search tools can perform the more routine actions of search, the "brainwork," in terms of cognitive and meta-cognitive skills, is still left to the user, and, in the case of inexperienced users, requires training to master. Studies of web queries submitted to commercial search engines have indicated that the use of operators is less important to query performance than term selection (Jansen, 2000; Lucas and Topi, 2002). This lends further support to the importance of training for the first two (offline) stages of the search process, in which articulation of the information need and conceptualization of the query occur. An interesting, but as of yet unanswered, question is the relationship between training in the specific domain of interest vs. general training in the process of articulating an information need within any domain.

At the same time, errors from a lack of semantic knowledge must not be ignored. Training for formulating an executable query (also part of Stage 2) and implementing that query in accordance with the rules governing the search system (Stage 3) can be critical to query success. Research supports the premise that training is required in order for searchers to be capable of writing correct Boolean queries (Ingwersen and Willet, 1995; Turtle, 1994). This is particularly true for systems in which queries require complex operator usage. In terms of search system usage, familiarity with the rules is not only helpful for Stage 3, but also comes into play in Stage 5, where search output must be correctly interpreted.

Finally, while technical errors are often common and affect performance in Stage 4 (entering the query), they are less critical to search performance than conceptual and semantic errors. Plus, many of these types of errors can and should be caught by the search tools, which are better suited to such tasks than the human operator.

5.3. Application of Results from Information Systems Training Research

Researchers in the field of Information Systems have spent a significant amount of time and effort during the past 15 to 20 years on end-user training in a variety of individual computing skills, such as the use of spreadsheets, word processing programs, web browsers, etc. Some of this training has focused on relatively simple technical skills (similar to those required to perform Stage 4 activities in the Figure 1 model), but other streams of this work have covered training that includes a much broader range of skills, including those requiring relatively complex modeling of the real-world problem domain (similar to Stage 1 activities in our model). In this section, we will briefly review some of the key findings from the relevant computer software training research that are potentially applicable to search training.

Since the seminal work by Compeau and Higgins (1995), IS research on computer training has widely utilized Bandura's (1986) Social Cognitive Theory (SCT) as the underlying model for understanding human behavior in training situations (Johnson and Marakas, 2000; Yi and Davis, 2001, 2003; Davis and Yi, 2004). Of particular note has been the identification of the importance of the self-efficacy construct and the selection of behavioral modeling as a foundation for designing training interventions. SCT, as a very complex general theory of human behavior, is far too comprehensive to review here in its entirety. Therefore, we will focus on the characteristics of the theory that IS researchers have found to be useful in computer skills training and evaluate their suitability for search training.

One of the strongest and most consistent findings in recent IS training literature has been the promise of behavior modeling as a training method, in comparison to many other training methods, such as traditional lectures, computer-aided instruction, and self-study using a paper-based manual (Yi and Davis, 2003). Compeau and Higgins (1995) reviewed the literature on a number of training fields and found that behavior modeling has been successful in a variety of those fields, such as mathematics, sales skills, and leadership. The core idea of modeling-based training is very simple: trainees develop their skills by seeing an expert perform a task successfully and using this experience as a basis for their own attempts.

The simplicity of this idea should not, however, be seen as a reason to reject the usefulness of the theory; on the contrary, SCT provides a strong rationale for understanding

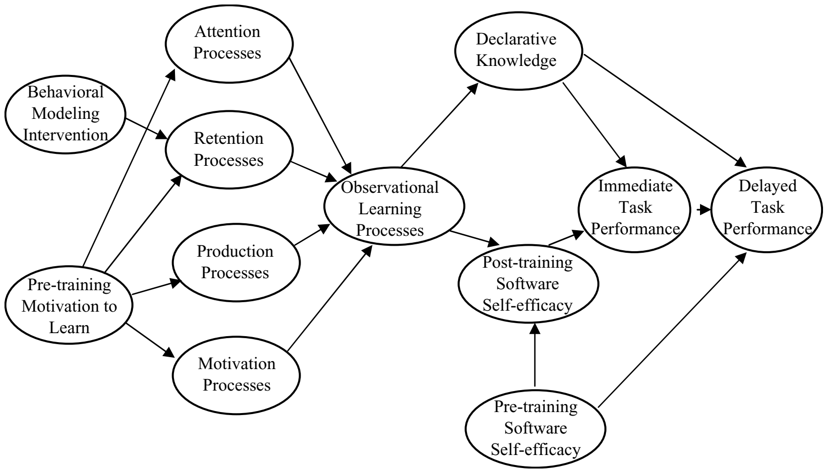

Figure 4: An example of a SCT-based training effectiveness model (Yi and Davis, 2003)

the underlying cognitive processes. As outlined by Yi and Davis (2003), SCT suggests that modeling-based training affects training outcomes through one or more of four observational learning processes: attention, retention, production, and motivation. In their empirical study on the effects of spreadsheet training, Yi and Davis (2003) found that the effects of a specific behavioral modeling intervention on learning outcomes were mediated by only one of the observational learning processes, retention. More important than this finding, or any other specific finding from this stream of research, is the use of a general psychological (learning) theory to explain why and how a specific training intervention may or may not affect the training outcomes.

Another very important benefit from the use of SCT in IS training literature has been the clarification of the role of self-efficacy in various learning processes. SCT proposes that self-efficacy, one's perceived capability to perform a specific activity (Kraiger, Ford, and Salas, 1993), significantly affects an individual's task performance and should be included in all models that are developed for understanding training interventions and learning behavior in specific situations. For example, Yi and Davis (2003) point out that self-efficacy as a construct is clearly distinct from declarative knowledge and has its own effects on immediate task performance. This was empirically verified in their study. Figure 4 presents the version of the spreadsheet training research model from Yi and Davis (2003) that was supported in their empirical study. We are not proposing that this or any other similar model would be directly applicable to search training. We do, however, believe that search training research could significantly benefit from similar efforts to utilize established psychological theories to model the mechanisms through which various training interventions affect search behaviors and outcomes. It appears that different stages in the search process lead to different cognitive processes and, therefore, would need different training models. Not enough is known about the specific characteristics of the mechanisms involved to enable the proposing of any detailed models at this time.

6. FUTURE DIRECTIONS

It is vitally important for the search training research community to continue the development of theoretically justified, conceptually sound models of the effects of various training interventions on learning processes during different stages of the search process. As shown earlier in this chapter, the extant search training research, for the most part, lacks focus and is not based on well-defined theoretical frameworks. Only first, relatively small steps have been taken towards a better understanding of the mechanisms that help end-users perform better searches. If future work in this area were based on well-defined conceptual models, it would be much more likely that results from separate research efforts could be integrated into and interpreted as a cohesive body of work.

We propose that the conceptual models developed in fields that also combine complex modeling-types of activities with computer technology usage could offer good starting points for these theory-development activities. For example, research on spreadsheet usage in information systems deals with questions that are similar to those of search research. Both require that the person performing a task understands the problem domain and the information need, is able to convert the information need into a conceptual model, expresses the model with an information technology solution, and interprets the results in a meaningful way. Therefore, it is likely that research on spreadsheet training could offer a useful direction for search research. The results from such related fields are not, of course, directly applicable to search, but they are an underutilized resource that could prove useful in the development of sound theoretical models for search research.

In addition to well-defined theoretical models, it appears that search training research would benefit from research projects that use multiple methods and approaches to study the same phenomenon. For example, a research project could design a training intervention based on theoretical models and experiences from existing training approaches and then evaluate it, first with controlled laboratory experiments, and then with field experiments and participant interviews. These activities are time consuming and often difficult, but necessary for the advancement of theory and practice in this important area.

We need cooperation with academic researchers, information search professionals, companies focused on providing web search as a business, and funding agencies to make large long-term studies using multiple methods possible. As was mentioned earlier, Halttunen's (2004) and his colleagues' systematic development and analysis of pedagogical approaches to training IR provide a valuable step forward and an example that the field can follow.

7. CONCLUSIONS

Developing and supporting human search capabilities is at least equally important as developing the capabilities of search engines. Existing academic research suggests that particularly inexperienced users searching the Web utilize only a modest subset of the capabilities that the tools offer and are weak at understanding their real information needs, articulating them in a way that allows for effective searches, and interpreting search results in the context of those needs. This clearly suggests that training users to become better searchers is a worthwhile effort, and that understanding what makes certain search interventions successful and others not is vitally important for enabling users to make effective use of their time.

Our review of existing search training literature revealed relatively broadly scattered efforts that, to a large extent, were not programmatic or focused on developing well-defined theoretical models. This makes it difficult to integrate findings from different studies into cohesive bodies of work that would support the development and evaluation of new training interventions. In this chapter, we propose three approaches to using existing areas of research as a theoretical foundation for future work. Namely: a) understanding the nature of search processes is very important if we want to understand search training, because different stages of the search process require different training approaches; b) a thorough analysis of search errors forms a solid foundation for training approaches that helps users avoid the common errors; and c) training research from related disciplines, such as information systems, can be used to introduce new theoretical perspectives and suggest models that are also potentially helpful in search training.

Search is pervasive in modern life and performing effective online searches is a fundamentally important literacy skill, not only for 21st century knowledge professionals but for every individual who relies on material available from online sources. Understanding how we, as a research community, can help people become effective searchers is, therefore, of critical importance.

REFERENCES

Abdulla, G., Liu, B., and Fox, E. A. (1998). Searching the World-Wide Web: Implications From Studying Different User Behavior. *Paper presented at the WebNet98 Conference,* Orlando, FL.

Allen, B. L. (2001). Boolean Browsing in an Information System: An Experimental Test. *Information Technology and Libraries,* 20(1), 12–20.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bates, M. (1985). An Exploratory Paradigm for Online Information Retrieval. *Paper Presented at the Sixth International Research Forum in Information Science (IRFIS 6),* Frascati, Italy.

Borgman, C. L. (1996). Why Are Online Catalogs Still Hard to Use? *Journal of the American Society for Information Science and Technology,* 47(7), 493–503.

Chan, H. C., Tan, B. C. Y, and Wei, K. K. (1999). Three Important Determinants of User Performance for Database Retrieval. *International Journal of Human-Computer Studies,* 51(5), 895–918.

Cheng, G. Y. T. (2003). Educational Workshop Improved Information-Seeking Skills, Knowledge, Attitudes and the Search Outcome of Hospital Clinicians: A Randomised Controlled Trial. *Health Information and Libraries Journal,* 20, 22–33.

Cloyd, B. and Spilker, B. (2000). Confirmation Bias in Tax Information Search: A Comparison of Law Students and Accounting Students. *Journal of the American Taxation Association,* 22, 60–71.

Colaric, S. M. (2003). Instruction for Web Searching: An Empirical Study. *College and Research Libraries,* 64(2), 111–122.

Compeau, D. and Higgins, C. (1995). Application of Social Cognitive Theory to Training for Computer Skills. *Information Systems Research,* 6(2), 144–176.

Cooper, W. (1988). Getting Beyond Boole. *Information Processing and Management,* 24(3), 243–248.

Davenport, T. H. and Prusak, L. (1993). Blow Up the Corporate Library. *International Journal of Information Management,* 13, 405–412.

Davis, F. D. and Yi, M. I. (2004). Improving Computer Skill Training: Behavior Modeling, Symbolic Mental Rehearsal, and the Role of Knowledge Structures. *Journal of Applied Psychology,* 89(3), 509–523.

Debowski, S. (2002). Developing Effective Electronic Information Seekers. *Australian Journal of Management,* 27, 21–30.

Detlor, B. (1999). Leveraging the Corporate Library Through Web User Training. *Library Management,* 20(7), 393.

Fallows, D. (2005). *Search Engine Users*. Washington, DC: Pew Reports.

Fallows, D., Rainie, L., and Mudd, G. (2004). *Data Memo: The Popularity and Importance of Search Engines*. Washington, DC: Pew Reports.

Ford, N., Miller, D., and Moss, N. (2001). The Role of Individual Differences in Internet Searching: An Empirical Study. *Journal of the American Society for Information Science and Technology,* 52(12), 1049–1066.

Halttunen, K. (2003a). Students' Conceptions of Information Retrieval: Implications for the Design of Learning Environments. *Library and Information Science Research,* 25(3), 307–332.

Halttunen, K. (2003b). Scaffolding Performance in IR Instruction: Exploring Learning Experiences and Performance in Two Learning Environments. *Journal of Information Science,* 29(5), 375–390.

Halttunen, K. (2004). *Two Information Retrieval Learning Environments: Their Design and Evaluation.* Tampere University.

Halttunen, K. and Järvelin, K. (2004). Assessing Learning Outcomes in Two Information Retrieval Learning Environments. *Information Processing and Management,* 41(4), 949–972.

Halttunen, K. and Sormunen, E. (2000). Learning Information Retrieval Through an Educational Game. *Education for Information,* 18(4), 289–311.

Hölscher, C. and Strube, G. (2000). Web Search Behavior of Internet Experts and Newbies. *The International Journal of Computer and Telecommunications Networking,* 33(1–6), 337–346.

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. and Willet, P. (1995). An Introduction to Algorithmic and Cognitive Approaches for Information Retrieval. *Libri,* 45, 160–177.

Jansen, B. J. (2000). The Effect of Query Complexity on Web Searching Results. *Information Research: An International Electronic Journal,* 6(1).

Jansen, B. J. and Pooch, U. W. (2001). A Review of Web Searching Studies and a Framework for Future Research. *Journal of the American Society of Information Science,* 52(3), 235–246.

Jansen, B., J. Spink, A., and Saracevic, T. (2000). Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web. *Information Processing and Management,* 36(2), 207–227.

Johnson, R. D. and Marakas, G. M. (2000). The Role of Behavioral Modeling in Computer Skills Acquisition: Toward Refinement of the Model. *Information Systems Research,* 11(4), 342–365.

Kim, K.-S. and Allen, B. (2002). Cognitive and Task Influences on Web Searching Behavior. *Journal of the American Society for Information Science and Technology,* 53(2), 109–119.

Kirkpatrick, D. L. (1998). *Evaluating Training Programs: The Four Levels*. San Francisco: Berrett-Koehler Publishers.

Kraiger, K. J. Ford, K. and Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Journal of Applied Psychology,* 78, 311–328.

Kuhlthau, C. C. (1993). *Seeking Meaning: A Process Approach to Library and Information Services*. Norwood, NJ: Ablex Publishing.

Lazonder, A. W. (2000). Exploring Novice Users' Training Needs in Searching Information on the WWW. *Journal of Computer Assisted Learning,* 16(4), 326–335.

Lucas, W. and Topi, H. (2002). Form and Function: The Impact of Query Term and Operator Usage on Web Search Results. *Journal of the American Society for Information Science and Technology,* 53(2), 95–108.

Lucas, W. and Topi, H. (2004). Training for Web Search: Will It Get You in Shape? *Journal of the American Society for Information Science and Technology,* 41(2), 383–403.

Marchionini, G. (1992). Interfaces for End-User Information Seeking. *Journal of the American Society for Information Science,* 43(2), 156–163.

Pharo, N. (2002). *The Search Situation and Transition Method Schema: A Tool for Analyzing Web Information Search Processes.* Unpublished Ph.D. thesis, University of Tampere, Tampere, Finland.

Shneiderman, B., Byrd, D., and Croft, W. B. (1997, January 1997). *Clarifying Search: A User-Interface Framework for Text Searches.* Retrieved January, 2000, from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html

Silverstein, C., Henzinger, M., Marais, J., and Moricz, M. (1999). Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum,* 33(1), 6–12.

Sormunen, E. and Pennanen, S. (2004). The Challenge of Automated Tutoring in Web- Based Learning Environments for Information Retrieval Instruction. *Information Research—An International Electronic Journal,* 9(2), paper 169 [Available at: http://InformationR.net/ir/9-2/paper169.html].

Spink, A. and Saracevic, T. (1997). Interactive Information Retrieval: Sources and Effectiveness of Search Terms During Mediated Online Searching. *Journal of the American Society for Information Science,* 48(8), 741–761.

Spink, A., Wolfram, D., Jansen, B. J., and Saracevic, T. (2001). Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science,* 53(2), 226–234.

Sutcliffe, A., G. Ennis, M., and Watkinson, S. J. (2000). Empirical Studies of End-user Information Searching. *Journal of the American Society for Information Science,* 51(13), 1211–1231.

Te’eni, D. and Feldman, R. (2001). Performance and Satisfaction in Adaptive Websites: An Experiment on Searches Within a Task-Adapted Website. *Journal of AIS*, 2(3).

Topi, H. and Lucas, W. (2005). Searching the Web: Operator Assistance Required. *Information Processing and Management,* 41(2), 383–403.

Turtle, H. (1994). Natural Language vs. Boolean Query Evaluation: A Comparison of Retrieval Performance. In *Proceedings of SIGIR 1994*.

Yi, M. I. and Davis, F. D. (2001). Improving Computer Training Effectiveness for Decision Technologies: Behavior Modeling and Retention Enhancement. *Decision Sciences,* 32(3), 521–545.

Yi, M. I. and Davis, F. D. (2003). Developing and Validating an Observational Learning Model of Computer Software Training and Skill Acquisition. *Information Systems Research,* 14(2), 146–169.

SECTION V

CONCLUSIONS

CHAPTER 12

CONCLUSION AND FURTHER RESEARCH

AMANDA SPINK
Queensland University of Technology

CHARLES COLE
McGill University

1. NEW DIRECTIONS IN COGNITIVE INFORMATION RETRIEVAL (CIR)

New Directions in Cognitive Information Retrieval (IR) gathers user or cognitive approaches to IR research into one volume. The group of researchers focus on a middle-ground perspective between system and user. They ask the question: What is the nexus between the wider context of why and how humans behave when seeking information and the technological and other constraints that determine the interaction between user and machine? These researchers' concern for the application of user/cognitive-oriented research to IR system design thus serves as a meeting ground linking computer scientists with their largely system performance concerns and the social science research that examines human information behavior in the wider context of how human perception and cognitive mechanisms function, and the work and social frameworks in which we live.

The researchers in this volume provide an in-depth revaluation of the concepts that form the basis of current IR retrieval system design. Current IR systems are in a certain sense based on design conceptualizations that view

- the user's role in the user-system interaction as an input and monitoring mechanism for system performance;
- the system's role in the user-system interaction as a data acquisition system, not an information retrieval system; and
- the central issue in the user-system interaction as the efficacy of the system's matching algorithms, matching the user request statement to representations of the document set contained in the system's database.

But the era of matching-focused approaches to interactive IR appears to be giving way to a concern for developing interactive systems to facilitate collaboration between users in the performance of their work and social tasks. There is room for cognitive approaches to interaction to break in here.

The cognitive approach researchers in this volume focus on IR interaction facilitating users finding and using information in the performance of the task for which the information is being sought. Perhaps the distinction between system and cognitive approach research is the difference between the two words "finding" (as in retrieving) and "use," but also between the two words "data" and "information." The contextualization

Amanda Spink and Charles Cole (eds.), *New Directions in Cognitive Information Retrieval.* 229–233.

of data retrieval into information finding and use in the user-system interaction is perhaps the road to turning the IR system from a data acquisition system into an information acquisition system for the user. The challenge cognitive approach researchers have given themselves is to work within the constraints of current and near-to-medium future technology to modify and upgrade the model of the user-system interaction utilized in the technology so that these systems truly become information retrieval systems.

As outlined in summary form in the Introduction chapter, the book divides the chapters into three sections: (1) CIR concepts, (2) CIR processes, and (3) CIR techniques. (An integration of all the chapters is available in Chapter 2, this volume, by Cole, Beheshti, Leide and Large.)

This last and concluding chapter of the book provides an overview of conclusions and further research for CIR based on the new directions discussed in the chapters in the book. We divide the new directions (with some overlapping) in the book into four sections:

- New Directions in Integration with Human Information Behavior
- New Directions in Cognitive Studies
- New Directions in Social and Organizational Studies
- New Directions in Relevance and Evaluation Studies

2. NEW DIRECTIONS: INTEGRATION WITH HUMAN INFORMATION BEHAVIOR

One of the key directions identified by chapter authors for CIR research is the integration of CIR within the broader human information behavior (HIB) framework which makes up the human information condition. An HIB perspective for CIR seeks to create a more holistic understanding of CIR that takes into account the HIB context in which human-IR system interaction takes place.

- In Chapter 6, Spink and Cole examine user multitasking when seeking and searching an IR system. New directions in research should ask: Does the user's information seeking stage (in the performance of the task for which information is being sought) affect the number and nature of multitasking searching? New Directions in research should also examine the interplay between information and non-information tasks, as well as beginning to conceptualize the so-called interruptive HIB behaviors users engage in during search session like "vanity searching." Such HIBs should be analyzed in terms of the role these behaviors may play in providing coordinating or switching mechanisms between individual search and seeking tasks during a multitasking search session.
- In Chapter 7, Vakkari and Järvelin highlight the need for a more integrated understanding of information searching and seeking. New directions in research should include studies that take into account the user's changing goals during task performance, and consequently the information searching and seeking it generates is a process, with the search topic, goals and the process itself gradually changing as the task performance proceeds.
- In Chapter 11, Lucas and Topi believe that despite recognition of the growing need for search training that supports the cognitive processes and learning styles of individual searchers, there has been relatively little response from the academic community to date. Different stages of the search process

require different training approaches. New directions in search training should follow the example of research on spreadsheet usage in information systems. There, after analyzing the information needs and problem domain of the targeted user group, the analysis was converted into conceptual models with a logical information technology solution built into the model.

3. NEW DIRECTIONS: COGNITIVE STUDIES

A key area for further research is the need for a greater understanding of the cognitive dynamics of human—IR system interaction.

- In Chapter 2, Cole, Beheshti, Leide and Large highlight the need for user-system interaction to stimulate the user into a selection state. New directions in research should focus on a thorough investigation of the nexus between the user's task and the role of information in the performance of that task in order to create task-specific system messages capable of aligning the user's cognitive state with information channels to potential information in the IR system's database.
- In Chapter 3, Larsen and Ingwersen extend understanding of the cognitive overlaps in the polyrepresentation continuum towards a more holistic and explicitly cognitive framework for understanding the processes involved in IR. New directions should examine simulation studies that test which methods would be appropriate for matching different representations of the user's cognitive space with document representations, applied to simulated work task situations. The exploration of all the possible permutations of controlled for variables is required to simulate all achievable combinations, taking into account the dependency of domains, media and representation styles.
- In Chapter 5, Ford calls for a greater understanding of knowledge need and knowledge behavior, which give a broader, more contextualized perspective than the current terms information need and information behavior. New directions in research should focus on the procedural mechanisms the user requires to create an "evidence map" to record the integration, cumulation, and disputation of—and other forms of interaction between—disparate information sources, formulated in terms of their own particular problems and perspectives. Ford gives the example of Ted Nelson's ZigZag hypertextual system as a new direction.
- In Chapter 10, Hook and Börner examine the cognitive benefits to the user of domain mappings of information spaces called Knowledge Diagram Visualizations (KDVs). These KDVs provide the user with a sort of cognitive scaffolding for organizing the information search, and conveying to the user the intellectual landscape of a particular domain and the specific neighborhood of interest to the user. New directions for research should center on the design of KDVs that facilitate user understanding of a domain by conveying to the user the schematic, geo-spatial, temporal, semantic, or social organization of the underlying domain, as well as facilitating user recall of search terms and recognition of future paths or directions during the search.

4. NEW DIRECTIONS: SOCIAL AND ORGANIZATIONAL STUDIES

Further research is needed that explores human interaction with IR systems at the social and organizational level.

- In Chapter 8, Beheshti, Bowler, Large and Nesset describe information seeking, search and construction for children in a learning situation, where social dynamics represent challenges to collaborative learning tasks. As well as at the cognitive level, organizational and social variables and perspectives are also important in the modeling of human IR system use. New directions focus on creating emphatic IR systems for the classroom setting that engage children in browsing as well as analytic searching for their classroom activities. An alternative IR system for children will therefore include a virtual reality environment based on a familiar metaphor designed for browsing, and avatars that are intelligent enough to provide context sensitive help and tutoring whilst displaying appropriate emotions. Such a system should be integrated with a conventional portal or search engine, to be used as an alternative tool for those children who prefer browsing rather than analytical strategies.
- In Chapter 2, Cole, Beheshti, Leide and Large describe a study of undergraduates researching an undergraduate essay. The social and organizational features of the essay require the undergraduate to put their learning on display to the instructor marking the essay. New directions focus on developing our understanding of the nexus between the IR system, the user and the informational requirements of the user's particular task in seeking information.
- In Chapter 7, Vakkari and Järvelin's analysis of the user's task performance as a primary variable in information seeking and search represents a gateway to efficiently including organizational and social variables in CIR-oriented research and design. New directions focus on enriching the IR research framework by introducing new concepts for representing actors and their typical domains and tasks, which would enhance the explanatory power of the framework. Also, future investigations, including both lab type investigation and other studies which consider actor, problem and task features, should take into account that because task performance, and consequently the information searching it generates, is a process, the search topic, goals and process gradually change as task performance proceeds.

5. NEW DIRECTIONS: RELEVANCE AND OTHER EVALUATION STUDIES

The concept and reality of human relevance judgments has been a major area of CIR research.

- In Chapter 4, Ruthven provides an overview of the CIR relevance research and proposes the need for a more integrated view. New directions in research should focus on making relevance a more natural concept by methodically mapping out how the balances and relationships between the searcher and tools they use for searching, particularly how the user gathers evidence about relevant information and makes inferences from that evidence, differently at different times of the user's research process, and differently for different purposes, depending on the tasks for which people choose to use IR systems.

- In Chapter 9, Kelly highlights the practical importance of relevance and relevance judgments. Kelly proposes that implicit feedback techniques are a major avenue of research to improve human—IR system interaction. New directions for research should focus on efforts needed to be made to fully understand how behaviors can be used for implicit relevance feedback.

6. CONCLUSION

At present no overall integrative framework exists for CIR, but an integrated approach–i.e., that distinguishes separate concepts and processes in information seeking and information search then attempts to create an integrated conceptualization of the user who is both searching and seeking information within the larger context of HIB–has the potential to yield a more holistic theoretical and cognitive understanding that will assist IR and Web system designers.

This book provides an overview of new directions in CIR research. The field of CIR is broad, international, interdisciplinary and dynamic with tremendous potential to impact the everyday lives of people in both developed and developing countries as they increasingly need to interact with IR systems. This book is not an exhaustive or historical discussion of all possible areas of important and new directions in CIR research. Information retrieval, in all its technical, cognitive and other respects continues to be an intractable research problem and research area. Our goal in producing this book was to stimulate the thinking of authors and readers alike.

BIOGRAPHIES OF AUTHORS

Jamshid Beheshti

Jamshid Beheshti has taught at the Graduate School of Library and Information Studies at McGill University for twenty years, where he was the Director for the past six years. Currently, he is the Associate Dean (Administration) of the Faculty of Education at McGill University. Dr. Beheshti is the principal investigator on a Social Sciences and Humanities Research Council of Canada (SSHRC) funded research on *A Virtual interface for children's Web portals*. He is a co-investigator on two other SSHRC funded research dealing with designing Web portals for children and designing visualization tools for undergraduate students. His publications have appeared in the *Journal of the American Society for Information Science and Technology*, *Information Processing & Management*, and *Education for Information* among other international journals. He has presented numerous papers in Canada, United States, Europe, and the Far East.

Katy Börner

Katy Börner is an Assistant Professor of Information Science in the School of Library and Information Science, Adjunct Assistant Professor in the School of Informatics, Core Faculty of Cognitive Science, Research Affiliate of the Biocomplexity Institute, Member of the Advanced Visualization Laboratory, and Founding Director of the new Cyberinfrastructure for Network Science Center at Indiana University.

Her research focuses on the development of visualization methodologies (knowledge domains, users of 3D virtual worlds, interfaces to digital libraries), networks and diffusion of knowledge, and the development of infrastructures for large scale scientific collaboration and computation.

She is the co-editor of a special issue of PNAS 101 (Suppl. 1) on 'Mapping Knowledge Domains' April 2004, a special issue on 'Collaborative Information Visualization Environments' in PRESENCE: Teleoperators and Virtual Environments, MIT Press, to appear in Feb. 2005, and the special issue on 'Information Visualization Interfaces for Retrieval and Analysis' in the Journal of Digital Libraries, to appear in 2005.

She is the recipient of many fellowships and awards, including Outstanding Junior Faculty Award, Pervasive Technology Laboratories Fellowship, SBC Fellow, NSF CAREER Award, and Trustees Teaching Award.

For more information on her research agenda, teaching, and other activities, visit: http://ella.slis.indiana.edu/~katy/

Leanne Bowler

Leanne Bowler is a doctoral student at McGill University's Graduate School of Library and Information Studies. Her research interests lie in the area of information behavior, with a focus on the cognitive and affective aspects of information seeking. She has a Master of Library Science and a Master of Education, both from McGill University. Prior to her doctoral studies, she worked as an information professional in a variety of settings, including public, school, and academic libraries, hospitals, and literacy organizations.

Charles Cole

Charles Cole is a Research Associate at the Graduate School of Library and Information Studies, McGill University, Montreal, and an information design consultant (Colemining, Inc.). At McGill, he is currently participating in a three year Social Science and Humanities Research Council (SSHRC) funded project. The project objective is to design IR system interface devices for undergraduate social science students researching their course essays when they have just selected their essay topic and are exploring the topic using an IR system. Findings have been published in the *Journal of the American Society for Information Science and Technology* (two forthcoming in 2005). The objective of the research is to produce a prototype interface for undergraduate history students researching their essay; the interface will facilitate the students' interaction with representations of document organization schemes, enabling them to form effective queries to IR systems.

Charles Cole holds a BA (History-Geography) and M.L.I.S from McGill University, a Ph.D. from the University of Sheffield, and held a two year SSHRC funded post-doctoral fellowship at Concordia University's Department of Communications. He has published over 35 refereed articles in such journals as the *Journal of the American Society for Information Science and Technology*, *Information Processing & Management*, *Library Trends, Library Quarterly* and *Library and Information Science Research*. Charles has co-edited (with Amanda Spink) two special topic journal issues: one on everyday life information seeking for *Library and Information Science Research*, and a two-issue special topic on information seeking for the *Journal of the American Society for Information Science and Technology*.

Nigel Ford

Nigel Ford is a Professor in Sheffield University's Department of Information Studies, which has achieved the maximum rating in all of the UK Research Assessment Exercises (RAEs) conducted to date. He has extensive research experience in information seeking, information behaviour, educational informatics and research methodologies. He has published a wide range of empirical and theoretical work in the form of refereed international journal papers, conference presentations and research reports, and has published 4 books on aspects of artificial intelligence programming. He has directed a number of funded research projects relating to information seeking, including the role of human individual differences in Web searching, and the development of Internet-based information skills in schools. He is currently directing a project investigating Web-based information searching by the general public. He is particularly interested in cognitive aspects of information behaviour, and an early interest in modelling cognitive processes using intelligent computation techniques has led to a continuing interest in the generation of models of information behaviour designed to be suitable for driving intelligent adaptive information systems. He is also particularly interested in research methodologies, and has published a number of critical theoretical papers adopting a cognitive psychological perspective on methodologies and practices within information science.

Peter A. Hook

Peter A. Hook is currently a doctoral student at Indiana University—Bloomington where he is a member of Dr. Katy Börner's Information Visualization Laboratory. He has a J.D. from the University of Kansas (1997) and a M.S.L.I.S. from The University of Illinois (2000). Prior to doctoral study, he was a law librarian for four years. His primary research focus is information visualization. Particular interests include the visualization of knowledge organization systems, concept mapping, and the spatial navigation of bibliographic data in which the underlying structural organization of the domain is conveyed to the user. Additional interests include social network theory, knowledge organization systems, and legal bibliometrics and informatics. Homepage: http://ella.slis.indiana.edu/~pahook.

Peter Ingwersen

Peter Ingwersen, Research Professor, Ph.D. in 1991. Head of Department of IR Theory, 1993–2000. Visiting scholar at the European Space Agency, Italy, 1980–84. Visiting Professor at Rutgers University, USA, 1987 and the Dept. of Information Studies, Tampere University, Finland, 1999–2002. Research areas: Interactive IR; Evaluation methods for work task-based IR; Informetrics-Scientometrics & Webometrics. He has published several books, and more than 70 journal articles and conference papers, in addition to editing work. Among his academic awards are the 1993 Jason Farradane, UK; the 1994 ASIS New Jersey Chapter, USA, Distinguished Lectureship Award, and, in 2003, the Distinguished Research Award, American Society for Information Science & Technology. In 2005 he received the Derek de Solla Price Medal for his informetric and webometric research by the International Society of Scientometrics and Informetrics. He is member of the editorial boards of five internationally leading journals in I&D and has carried out a number of tutorials (Introduction to IR, with Peter Willett) at ACM-SIGIR conferences at the end of the 1990s.

Kalervo Järvelin

Kal Järvelin has been Professor of Information Studies (http://www.info.uta.fi/), University of Tampere, since 1996 and Academy Professor (the Academy of Finland) 2004–2009. He holds a PhD in Information Studies from the University of Tampere. Kal Järvelin's research covers information seeking and retrieval, and database management, especially structured documents; linguistic and conceptual methods in IR. He has authored and co-authored some 200 academic papers and articles.

Kal Järvelin has served on the ACM SIGIR Conferences as a program committee member (1992–2005), Conference Chair (2002) and Program Chair (2004, 2006). He is an editorial board member of Information Processing and Management (USA). For further information see Kal's personal home page at http://www.uta.fi/~likaja/.

Diane Kelly

Diane Kelly is an Assistant Professor at the School of Information and Library Science and the University of North Carolina at Chapel Hill. Her research interests include interactive information retrieval, user modeling, relevance feedback, personalization, information-seeking behavior, and evaluation methods and metrics. Chapter 9 in the present volume is based on one of her current research projects that seeks to identify and evaluate the use of online information-seeking behaviors as implicit relevance feedback for user modeling and personalized information delivery. In this work, Kelly is further concerned with understanding how contextual variables, such as specific task and topic, affect the use of behavior as implicit feedback, and how these variables can be measured and tracked over time. In addition to this project, Kelly has conducted research on tools that support interactive information retrieval at the search interface, such as those used for explicit relevance feedback and query expansion. Kelly has participated in TREC for seven years, and has also participated in research on the development of evaluation metrics for question-answering systems. Kelly serves on the Editorial Board of Information Processing and Management, and is active in ACM SIGIR, JCDL, and CoLIS. Kelly holds a bachelor's degree in Psychology and English from the University of Alabama, a M.L.S. and Ph.D. from Rutgers University and a Graduate Certificate in Cognitive Science from the Rutgers University Center for Cognitive Science.

Andrew Large

Andrew Large is CN-Pratt-Grinstad Professor of Information Studies at McGill University and former Director of its Graduate School of Library and Information Studies. His research interests encompass information-seeking behavior, information retrieval and human-computer interaction (HCI), especially as they involve children. He has published many articles, chapters and books in these fields, among his latest being *Digital Libraries: Principles and Practice in a Global Environment* (K.G. Saur, 2005), and the chapter on "Children, Teenagers and the Web" in the latest edition of the *Annual Review of Information Science and Technology*. He also edits the quarterly journal, *Education for Information*. Dr. Large has presented papers in more than twenty countries around the globe on various topics in information studies.

Birger Larsen

Birger Larsen is Assistant Professor at the Department of Information Studies at Royal School of Library and Information Science, Copenhagen, Denmark. He received his PhD in 2004 on the topic of using references and citations in automatic indexing and retrieval systems, proposing a new method, the boomerang effect, which allows best match searching via a citation network without known seed documents. His research interests include information retrieval (IR) and information seeking in structured documents, the use of citation indexes in IR and bibliometric research evaluation. He is a co-organiser of the interactive track at the Initiative for the Evaluation of XML Retrieval (INEX).

John E. Leide

John E. Leide is an Associate Professor in McGill University's Graduate School of Library and Information Studies. He obtained a BS (mathematics-humanities) from the Massachusetts Institute of Technology, a MS (Library Science) from the University of Wisconsin (Madison) and a Ph.D. (Library Service from Rutgers University. At the Graduate School of Library and Information Studies, John teaches organization of information, classification and cataloguing, systems thinking, and abstracting and indexing. John's current research interests are iintegrating classification visualization devices for undergraduate users of information retrieval systems, and bilingual thesaurus development in gerontology and geriatrics.

John Leide is currently Principal Investigator of a three year Social Science and Humanities Research Council (SSHRC) funded project, the objective of which is to design IR system interface devices for undergraduate social science students researching their course essays when they have just selected their essay topic and are exploring the topic using an IR system. Findings have been published in the *Journal of the American Society for Information Science and Technology* (two forthcoming in 2005). The objective of the research is to produce a prototype interface for undergraduate history students researching their essay; the interface will facilitate the students' interaction with representations of document organization schemes, enabling them to form effective queries to IR systems.

Wendy Lucas

Wendy Lucas is an associate professor in the Department of Computer Information Systems at Bentley College in Waltham, MA. She received her Ph.D. from the Electrical Engineering and Computer Science Department at Tufts University. Dr. Lucas was formerly president and cofounder of a software company specializing in medical office systems and has served as a marketing and technology consultant to small businesses. Her research interests include Web search, information retrieval and presentation, interface usability, and data visualization. She has published in the *Journal of the American Society for Information Science and Technology*, *Communications of the Association for Information Systems*, *Information Processing & Management*, and *Theory and Practice of Object Systems*. Dr. Lucas's primary teaching interests are programming languages, algorithms, object-oriented technologies, and business-related IT applications.

Valerie Nesset

Valerie Nesset is a PhD student at Graduate School of Library and Information Studies, McGill University. Her research interest is the information-seeking behaviour of younger elementary school students. She has co-authored a number of research papers and articles as a member of an Social Sciences and Humanities Research Council of Canada funded research project on Children and Web Portal Design.

Ian Ruthven

Ian Ruthven is a lecturer in the Department of Computer Science and Information Sciences at the University of Strathclyde, a position he has held since November 2001. His research interests include the modelling of interactive information retrieval systems and interfaces, contextual information retrieval and user-centred evaluation methodologies for information retrieval. He has collaborated in, and led, a number of national and international research projects including projects funded by the European Union, British Library Research Council, British Engineering and Physical Science Research Council and the British Royal Society. Ruthven holds a PhD in interactive information retrieval from the University of Glasgow and a Master's degree in Cognitive Science from the University of Birmingham.

Amanda Spink

Amanda Spink is Professor of Information Technology at the Faculty of Information Technology, Queensland University of Technology, Brisbane, Australia. She has a B.A. (Australian National University); Graduate Diploma of Librarianship (University of New South Wales); M.B.A. in Information Technology Management (Fordham University), and a Ph.D. in Information Science (Rutgers University). Amanda's research focuses on theoretical and applied studies of human information behavior and interactive information retrieval (IR), including Web and digital libraries studies, and information science theory. The National Science Foundation, the American Library Association, Andrew R. Mellon Foundation, Amazon.com, NEC, IBM, Excite.com, AlltheWeb.com, AltaVista.com, FAST, and Lockheed Martin have sponsored her research. She currently has two large-scale research projects underway with the Web companies Vivisimo.com and InfoSpace, Inc.

Amanda has published over 220 journal articles, refereed conference papers and book chapters, with many in the *Journal of the American Society for Information Science and Technology*, *Information Processing and Management*, *Journal of Documentation*, *Library Trends*, *Interacting with Computers*, *IEEE Computer*, *International Journal of Information Management*, *Information Research*, *Online Information Review*, *Cyberpsychology and Behavior*, *Internet Research*, *the ASIST*, *IEEE ITCC*, *CAIS*, *Internet Computing*, *ACM SIGIR*, and *ISIC Conferences*. Her new book Web Search: Public Searching of the Web, published by Springer, is the first research monograph detailing human interaction with Web search engines. Amanda co-edited with Dr. Charles Cole (McGill University), the books New Directions in Cognitive Information Retrieval and New Directions in Human Information Behavior—Springer.

Heikki Topi

Dr. Heikki Topi is currently an Associate Professor of Computer Information Systems at Bentley College in Waltham, MA and Associate Director of Bentley's MSIT program. He received his Ph.D. in Management Information Systems from the Graduate School of Business at Indiana University. His teaching interests cover a range of topics, including advanced systems analysis and design, data management, data communications, and telecommunications and information industries. His current research focuses on human factors and usability issues in information search and data management, management and commercial utilization of advanced telecommunications technologies with a special emphasis on wireless applications, and the effects of time availability on human-computer interaction. His research has been published or will appear in the *International Journal of Human-Computer Studies, Journal of the American Society for Information Science and Technology, Information Processing & Management, Small Group Research, Journal of Database Management,* and*Communications of the Association for Information Systems.* He is the co-editor of Auerbach's *IS Management Handbook, 8th Edition* with Carol V. Brown. He also served on the ACM/AIS/AITP joint task force that developed *IS 2002*, the newest undergraduate IS model curriculum, and is currently a member of a joint ACM/AIS/IEEE-CS task force working on *CS2005 Overview Report.*

Pertti Vakkari

Dr. Pertti Vakkari is Professor of Information Studies at the University of Tampere, Finland. He has been the chairperson of the Finnish Doctoral Programme for Information and Communication Studies in 1994–2002. He has served as chairperson in 1998–2002, and serves as board member in 2004–2007 in the Nordic Information Studies Research Education Network.

He has been a member of several program committees of international conferences in information studies like CoLIS, ISIC and SIGIR. He is a member of the editorial board of journals *Information Processing & Management*, *Archival Science and Journal of Documentation*.

Prof. Vakkari's research activities may be broadly divided into 1) studies on information seeking, 2) studies on relations between task performance and information search process and outcomes 3) studies on interactive query expansion based on relevance feedback and thesaurus , and 4) theoretical and empirical analyses of information science.

His publications include several monographs and readers and over 100 research articles.

INDEX