SUPPORTING MULTIPLE INFORMATION-SEEKING STRATEGIES IN A SINGLE SYSTEM FRAMEWORK

by

XIAOJUN YUAN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Communication, Information, and Library Studies

written under the direction of

Nicholas J. Belkin

and approved by

____

____

____

____

____

New Brunswick, New Jersey

October, 2007

ABSTRACT OF THE DISSERTATION

SUPPORTING MULTIPLE INFORMATION-SEEKING STRATEGIES

IN A SINGLE SYSTEM FRAMEWORK

By XIAOJUN YUAN

Dissertation Director:
Nicholas J. Belkin

This study explores issues in information retrieval (IR) systems with special attention to information-seeking strategies (ISSs), the relation of ISSs to IR system design, and how to support multiple ISSs within a single system framework. It addresses the observation that people engage in a variety of ISSs within a single information-seeking episode. This study proposes to construct and evaluate an interactive IR (IIR) system which incorporates different IR support techniques to adaptively support multiple ISSs. Based on an information-seeking episode model (Belkin, 1996), and a multi-faceted classification scheme of information behaviors (Cool & Belkin, 2002), it was conducted in a series of three consecutive steps. Firstly, four experimental systems were designed and implemented with each tailored to one of the following IR support techniques: database summary, clustered retrieval results, table of contents navigation, and fielded query. A within-subjects experiment was conducted to compare each experimental system to its respective generic baseline system, which was constructed by following the current standard model with a specific query input and a ranked list of search results. Results indicated that the experimental systems were superior to the baseline systems. Secondly, information-seeking dialogue structures developed in the MERIT system (Belkin, Cool, Stein & Thiel, 1995) were adopted to guide the design of the IIR system. The dialogue structures were built based on the Conversational Roles (COR) model (Sitter & Stein, 1992). Finally, an experimental system which supported multiple ISSs was built by incorporating the four IR support techniques and the dialogue structures. This experimental system was tested in a within-subjects experiment in comparison to a generic baseline system. The experiment, with 32 subjects each searching on eight different topics, indicated that using the experimental system resulted in significantly better performance, significantly more effective interaction, and significantly better usability than the baseline system. These results demonstrated that it is possible to support quite different information-seeking behaviors within a single system framework which searchers can understand and use effectively. A principled approach to designing such systems needs to be further investigated.

ACKNOWLEDGEMENTS

My long journey to complete my doctoral dissertation would not have been possible without the support of many wonderful people, including my advisor, my committee members, my colleagues, my friends, and my family.

I would like to express my deep gratitude to my dissertation advisor, Dr. Nicholas J. Belkin, for his unparalleled insight, guidance, patience, and uncompromising emphasis on quality and meaningful research.

I would also like to thank my committee members, Dr. Susan Dumais, Dr. Michael Lesk, Dr. Anselm Spoerri, and Dr. Chengxiang Zhai. They are all very outstanding researchers and mentors. What I have learned from them will benefit me for my lifetime.

Many thanks go to researchers in SCILS, Dr. Tefko Saracevic, Dr. Nina Wacholder, and Dr. Xiangmin Zhang, with whom I have worked on various research projects. I would also like to thank Dr. Daniel O'Connor, Dr. Claire McInerney, and Dr. Marie Radford, who gave me invaluable suggestions on my career path.

There are researchers outside of SCILS that I would like to thank for their insightful comments on my dissertation at various doctoral consortiums (ASIST'04, SIGIR'04, HLT-NAACL'06): Macia Bates, Alan Black, Ciprian Chelba, Bruce Croft, David Harper, Caroline Haythornthwaite, Ed Hovy, Barbara Kwasnik, Liz Liddy, Yoëlle Maarek, Alistair Moffat, Javed Mostafa, Doug Oard, Keith van Rijsbergen, Steve Robertson, Linda Schamber, Henry Small, Paul Solomon and John Tait. I would also like to thank the anonymous reviewers from the SIGIR'07 conference.

I would like to thank each agency which has provided financial support for my research: NSF Grant #99-11942, and the Eugene Garfield Doctoral Dissertation Fellowship.

I would like to thank the researchers who assisted me with the technical suggestions and support, including David Fisher from University of Massachusetts at Amherst, Paul Ogilvie from Carnegie Mellon University, and Liang Zhou from University of Southern California. I would also like to thank Jon Oliver for his technical support over the years.

I want to thank my friends in Rutgers. During the past several years, their friendship helped me overcome a lot of difficulties, either in study or in life.

I would like to thank my parents, my husband, and my other family members. Their selfless support is something I can always count on. We have gone though many joyful and sad moments together during my years of study. Without them, I would not be where I stand right now.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Chapter 1

Introduction

Standard, traditional information retrieval (IR) systems are designed to support only one information-seeking strategy (ISS): specified searching (Belkin, 1993). The underlying assumption is that people are usually clear about what their information need is. However, this is not always the case. Most times, people do not know exactly which terms they should employ to achieve desired results, especially when they are not familiar with the system or they only have a vague idea about what they are looking for. In the former case, people might want to get familiar with the new system first before they start a "real" search. They might want to take a look at the system by browsing the structure, finding out the coverage of the contents, then specifying queries for retrieving documents they want. In the latter case, people might want to learn from the system first for the purpose of formulating good queries. They might do a search first, then browse the retrieved results to see whether some documents stimulate their recollection or lead to a better formulated query in order to get what they want.

Current research studies have identified other issues ignored by traditional IR systems.

Firstly, human information-seeking behavior is much more complicated than just query formulation and term selection. It includes getting an idea of which domain or genre of information people need, familiarizing themselves with the content and structure of a variety of databases, learning about a domain of interest, extending their knowledge of this domain in order to formulate more effective queries, changing their searching strategies to improve their queries, and other behaviors.

Secondly, human information-seeking behaviors are not discrete processes. These behaviors interact with one another during information-seeking processes (Belkin et al., 1990; Lin & Belkin, 2000). For example, a person who is interested in discovering some comments from an electronic book might enter the system with a search for a specified book. If this book is available, the person might look through the table of contents in order to find possibly related comments.

Thirdly, interaction, not comparison or representation, is the central process of IR (Croft & Thompson, 1987; Belkin, 1993). Belkin (1993) further suggested that IR should be considered as an inherently interactive process, and that IR systems should be designed to support appropriate interactions.

Fourthly, people with different goals and tasks need to use different information-seeking strategies (ISSs) to conduct searches. For example, a person who is interested in finding some relevant documents on global warming might first look at the overall structure of the system, get familiar with the databases in the system, then enter a database by specifying queries to select relevant documents. However, traditional IR systems ignore the fact that people employ a variety of ISSs during their information-seeking processes. Information-seeking behavior is characterized by movement from one ISS to another within a single information-seeking episode (Belkin, Marchetti & Cool, 1993). An information-seeking episode can be viewed as a series of interactions between the person and information objects (Belkin, 1996).

Lastly, to build an IR system which can support all ISSs, it is necessary to employ different combinations of IR support techniques. Different IR systems can optimally support different ISSs, such as browsing or specified searching (Oddy, 1977; Bates, 1990). These different kinds of ISSs can be optimally supported by different combinations of IR support techniques (Belkin, 1996). Some examples of IR support techniques are: for *comparison,* best match; for *navigation*, following links; for *representation*, indexing; for *visualization*, lists. The diversity of ISSs indicates that an IR system supporting one ISS well is unlikely to support the others at the same level. Conversely, an IR system which employs only one combination of techniques to support all ISSs is unlikely to be successful at best support for any one of them.

Therefore, this study proposes to design an interactive IR (IIR) system which incorporates different IR support techniques to *adaptively* support different ISSs. In other words, an IIR system should be able to provide different combinations of IR support techniques for different ISSs, by making support available for different ISSs, or by recognizing when a searcher is likely to engage in a specific ISS, and providing support for that ISS. This study takes the former approach.

Although there have been some attempts to design systems (or to propose frameworks for systems) which will support more than one ISS (Belkin et al., 1993; Olston & Chi, 2003), for the most part this issue has been ignored in both research and operational IIR systems. It seems that this situation is due predominantly to three factors: the lack of recognition of the problem itself; the lack of a theoretical structure which might provide a framework within which multiple ISSs could be supported; and, the inherent difficulty of the task itself.

Our solution is illustrated in Figure 1.1. In this study, the main focus will be on parts 2 and 3, based on existing theoretical and empirical findings with respect to part 1.

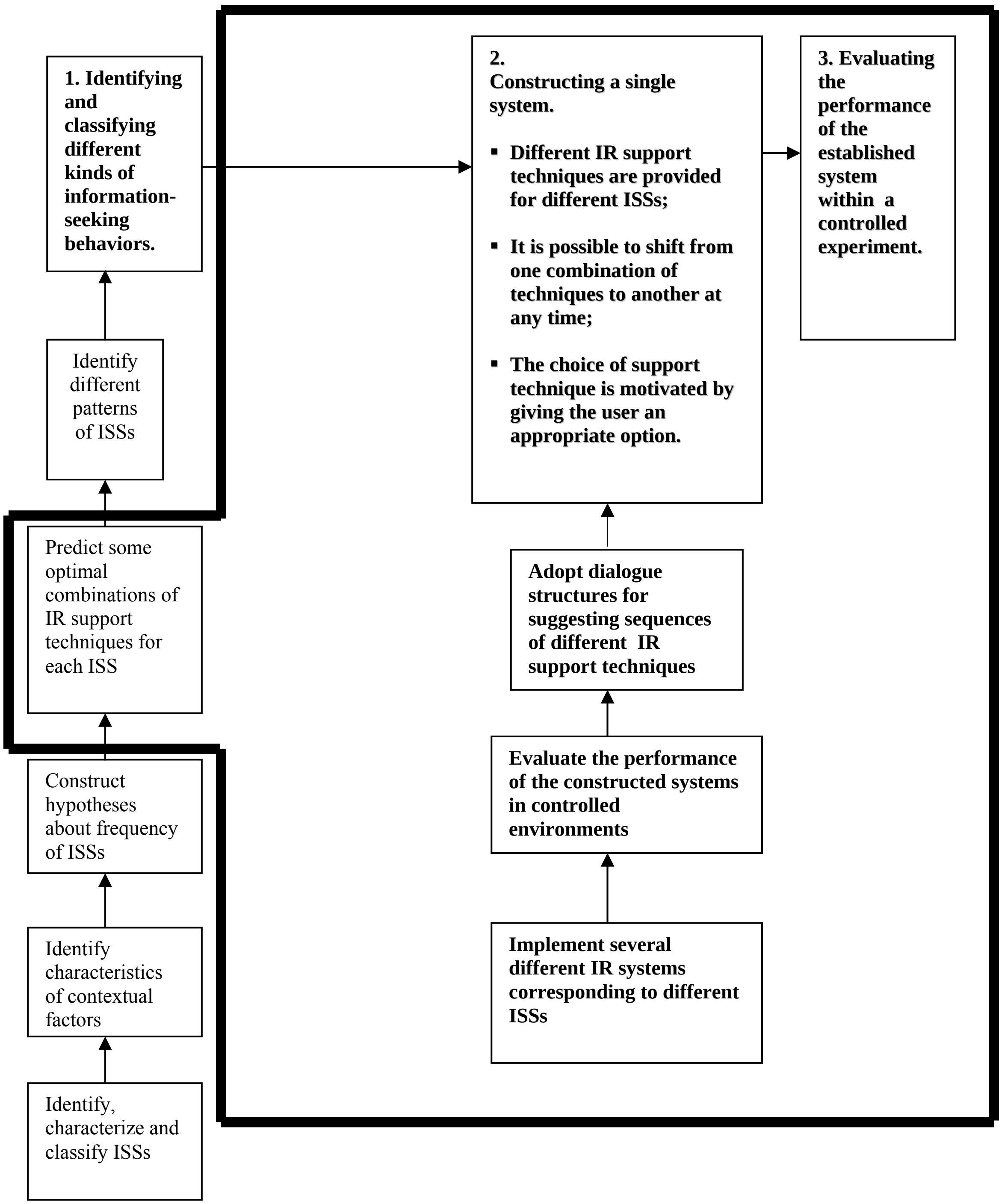


Figure 1.1 Research problems

This study is mainly concerned with building an IIR system which adaptively supports multiple ISSs during the information-seeking process. By *adaptively* it means that support for different ISSs is available to the searcher within the single IIR system. To achieve this goal, the following research problems were investigated:

Research problem 1:

Implementing and evaluating several systems which were tailored to different ISSs. These systems were based on different IR support techniques. Implementation of the systems employed the LEMUR Toolkit (http://www.cs.cmu.edu/~lemur) and hypotheses about the optimal combination of IR support techniques. Evaluation was done in controlled experiments using the metrics developed in the interactive track of TREC (Text REtrieval Conference).

Research problem 2:

Adopting dialogue structures for guiding and controlling sequences of different IR support techniques. The dialogue structures adopted are the ones used in the MERIT system (Belkin et al., 1995).

Research problem 3:

Constructing and evaluating an IIR system which adaptively supports various ISSs. This system was implemented based on the evaluation results of research problem 1 and the dialogue structures in research problem 2. Evaluation was conducted in a controlled experiment using the metrics developed in the interactive track of TREC.

Chapter 2

Literature Review

This study brings together such areas as information-seeking behavior, ISSs, information-seeking dialogues, and integrated IR systems. This section reviews previous studies in these areas. It begins with an overview of information-seeking behavior models. This is followed by a description of studies on ISSs, with special attention to the multi-dimensional classification schemes of ISSs. It then discusses the models of information-seeking dialogues, and, finally, the problems of integrated IR systems.

2.1 Information-Seeking Behavior Models

Studies on information-seeking behavior have tried to establish models which can identify information-seeking patterns and some contextual variables such as information-seeking stages.

Based on the observation of information-seeking behaviors of library users, Belkin et al. (1990) identified a pattern of information-seeking behaviors; that is, initiating a specific search statement, browsing when reaching the shelves with the located items, evaluating items based on their usefulness for the information problem, and employing citations of these items to search for other interesting items.

Ellis (1989) and Ellis, Cox and Hall (1993) developed a behavioral model of information-seeking. In his model, Ellis discovered eight different activities: *starting* (beginning to look for information), *chaining* (following citations), *browsing*, *differentiating* (filtering information), *monitoring* (keeping up-to-date), *extracting* (finding relevant material), *verifying* (checking information accuracy) and *ending*. Ellis's model does not define any order between the different activities.

Kuhlthau's Information Search Process (ISP) model (Kuhlthau, 1991) made up for some shortcomings of Ellis's model by relating associated feelings, thoughts and actions to the information-seeking stages, and by providing some sequential structure. Kuhlthau's model identified six successive stages, each of which is in turn connected with a specific activity in the information-seeking process: *initiation* (recognize), *selection* (identify), *exploration* (investigate), *formulation* (formulate), *collection* (gather) and *presentation* (complete). Table 2.1 shows, for example, how the Initiation phase is associated with feelings of uncertainty, general or vague thoughts about the information problem, and the action of seeking background information. The appropriate task here is to recognize the information problem.

These traditional information-seeking behavior models are helpful for describing how people conduct their searches during an information-seeking episode, but it is still not clear how their information-seeking behaviors would be supported during their interaction with IR systems.

Table 2.1

Information Search Process (ISP) (after Kuhlthau, 1991)

Stages in ISP	Feelings common to each stage	Thoughts common to each stage	Actions common to each stage	Appropriate task according to Kuhlthau model
1. Initiation	Uncertainty	General/ Vague	Seeking background information	Recognize
2. Selection	Optimism			Identify
3. Exploration	Confusion/ Frustration/ Doubt		Seeking relevant information	Investigate
4. Formulation	Clarity	Narrowed/ Clearer		Formulate
5. Collection	Sense of direction/ Confidence	Increased interest	Seeking relevant or focused information	Gather
6. Presentation	Relief/ Satisfaction or disappointment	Clearer or focused		Complete

2.2 IIR Models

The traditional IR model (see Figure 2.1) represents IR simply as elements and processes related to system and user, with primary focus on the system side. As can be seen from Figure 2.1, the central process is the comparison of query and text surrogates, which relies on the two representation processes. This model places most emphasis on text representation, with relatively little attention on representation of information need. This model is strong in that it allows the comparison of diverse methods and algorithms by employing common evaluation methods. However, it does not address interaction in IR and encounters many difficulties in evaluating interactive aspects in IR. Thus the need for research in an IIR model arises.

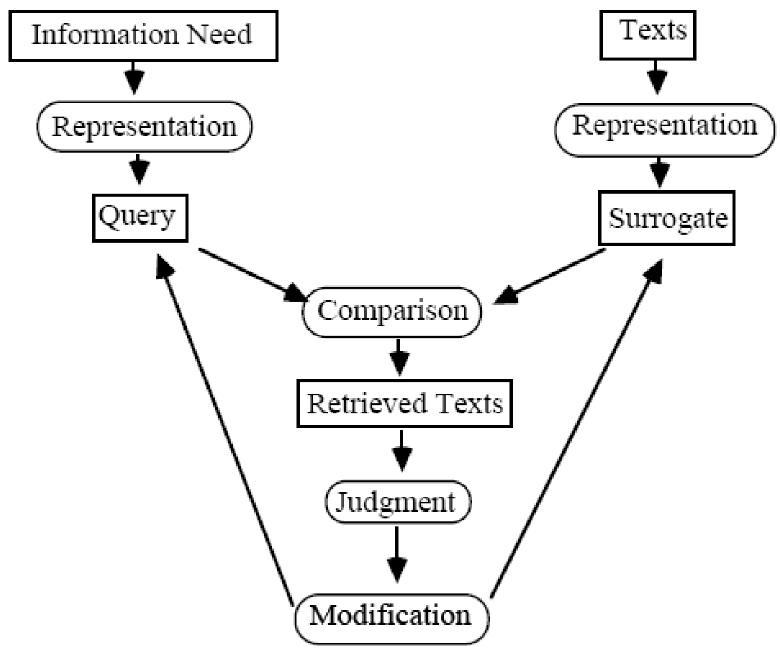


Figure 2.1 Traditional IR model (after Belkin, 1993)

Generally speaking, an IIR model is a model that takes into account the interactive, cognitive, affective and situational aspects of IR, and concentrates on the interaction between the user and the IR system or information objects.

Ingwersen's cognitive model (Ingwersen, 1996) (Figure 2.2) proposes that interaction in IR is a set of cognitive processes which occur in all the information processing elements in IR, and that users interact with both IR systems and information objects. These interactions occur at different levels with different types. This model also addresses the fact that a polyrepresentation is applied to both the user's cognitive space and the IR systems' information space.

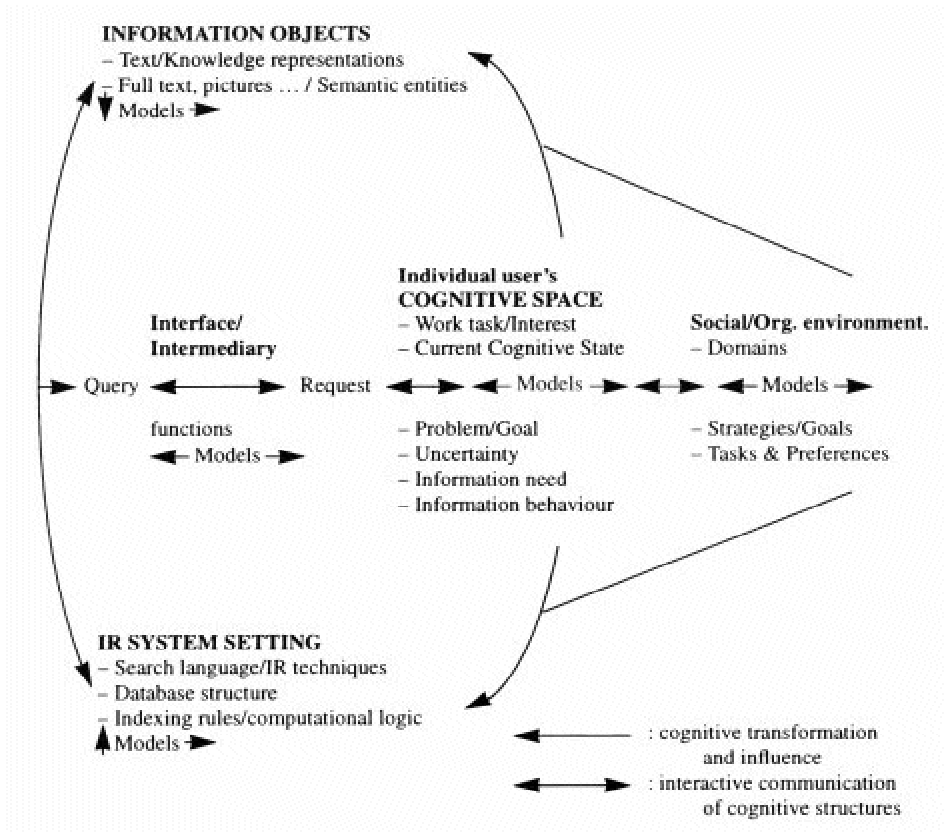


Figure 2.2 Ingwersen's cognitive model (after Ingwersen, 1996)

Saracevic's stratified model (Saracevic, 1996) (see Figure 2.3) assumes that users interact with IR systems for the purpose of using information and that such use is related to cognition and situational application. This model is composed of two major elements, user and computer, each with several different levels. The user side has four levels: surface level (users conduct dialogues with computers through an interface), cognitive level (users interact with information objects), affective level (users interact with their affections), and situational level (users interact with the situation). The computer side has also four levels: surface level, engineering level (focuses on effects of hardware attributes), processing level (stresses on algorithms), and content level (concentrates on contents of information sources). A series of adaptations could occur on both the user and the computer side, and move toward the surface level. Interaction is considered as a series of processes which occur in several related levels. This model decomposes various interactions into different elements and stresses the cognitive, affective and situational aspects of IR interaction.

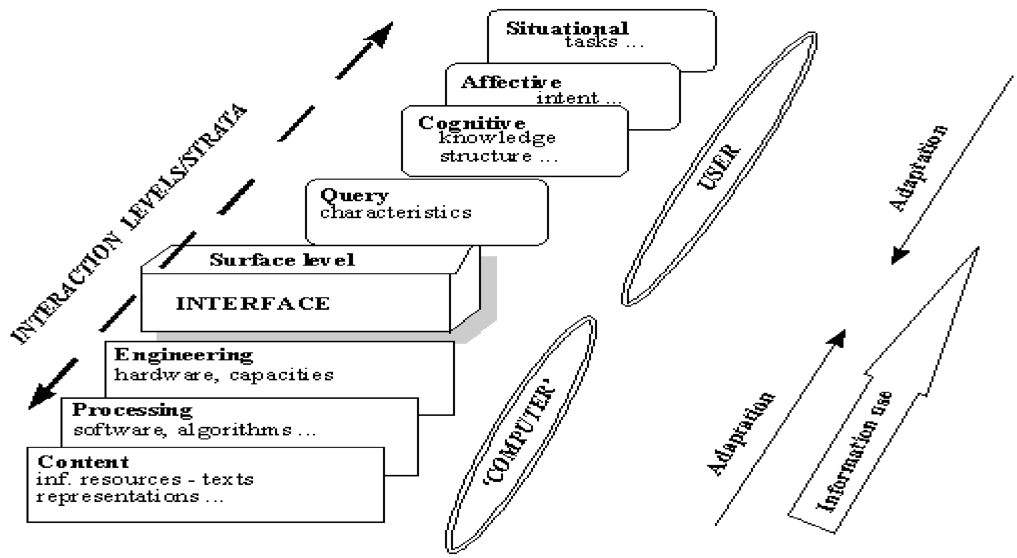


Figure 2.3 Stratified model of IR interaction (after Saracevic, 1996)

Spink's model (Spink, 1997) (see Figure 2.4) is an extension of the two models just discussed, with more focus on the nature and role of feedback in IIR. In her model, an interactive search process is composed of a series of search strategies which consist of several cycles, with several interactive feedback loops in each cycle. An interactive feedback is made up of multiple search tactics or moves, and user judgments. The strength of this model is that it includes the user in the feedback loop, and it recognizes varieties of interactive feedback which provide the communication between the user and the IR system. Also this model addresses the situational and cognitive state of the user.

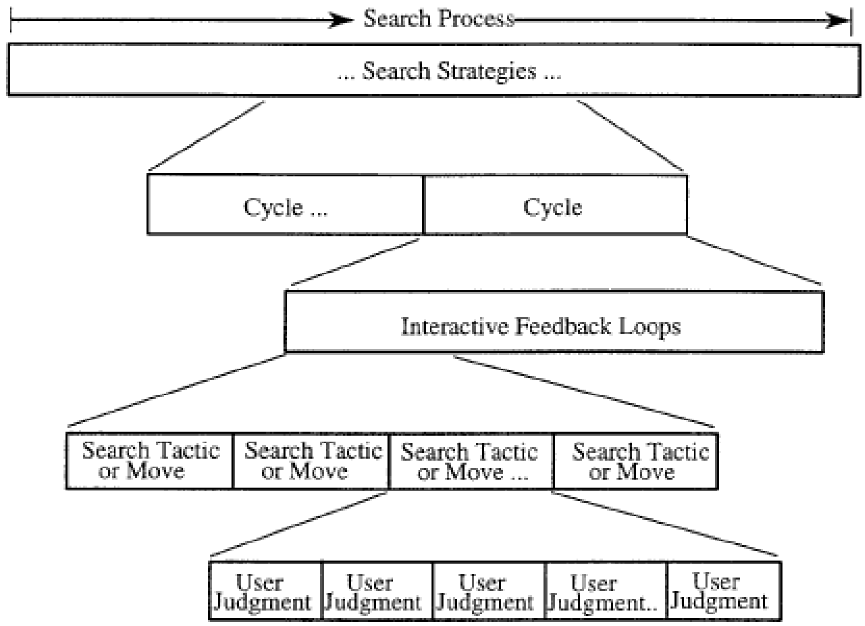


Figure 2.4 Elements of the interactive process (after Spink, 1997)

The strength of the above models is that they focus on the cognitive, situational, and interactive aspects of IR. But they do not account for the variety of IR processes.

Belkin's information episode model (Belkin, 1996) (see Figure 3.1) assumes that representing an Anomalous State of Knowledge (ASK) or information need is the central problem in IR. Different from the models discussed above, this model not only takes into account cognitive and situational aspects, but also directly addresses a number of IR processes: comparison, summarization, representation, visualization, and navigation. Each of these processes can be enumerated in different ways. Interactions between the user and the IR system depend on a number of factors, such as goals, intentions and tasks. The varieties of interactions include processes such as judgment, interpretation, and modification. The strength of this model lies in describing ISSs in more detail and suggesting IR support techniques which can optimize these ISSs.

In summary, the traditional IR model does not concentrate on interaction. The interaction models discussed so far depict interaction in different ways with various foci. All these models need to be brought into practical applications.

2.3 ISSs

Studies in this area have tried to identify the different varieties of moves, tactics or ISSs and explore the reasons that stimulate people to employ different ISSs.

Fidel (1985) discovered two moves that modify query formulations: *operational moves*, and *conceptual moves*. These moves allow users to view possible options of query formulation modification.

Bates (1979) identified 29 information tactics and grouped them into four categories: *monitoring tactics* (making the search on track), *file structure tactics* (navigating through the file structure), *search formulation tactics* (creating a search formulation) and *term tactics* (selecting and changing terms).

Focusing on online search, Harter and Roger-Peters (1985) grouped 101 tactics into six categories: *overall philosophical attitude and approach, language of problem description, record and file structure, concept formulation and reformulation, recall and precision, and cost/efficiency*.

Shute and Smith (1993) identified 13 knowledge-based search tactics and classified them in terms of their effect on topic scope: *broaden topic scope*, *narrow topic scope* and *change topic scope*.

Marchionini (1995) classified ISSs into two high-level groups: *analytical strategies*, which are more goal-oriented and systematic, and *browsing strategies*, which are more interactive and informal. Popular analytical strategies include *building bocks*

(Harter, 1986), *successive fractions* (Meadow & Cochrane, 1981), *pearl growing* (Markey & Cochrane, 1981), and *interactive scanning* (Hawkins & Wagers, 1982). Browsing strategies include *scanning*, *observing*, *navigating* and *monitoring* (Marchionini, 1995).

Belkin, Cool, Koennman, Ng, and Park (1996) proposed a taxonomy of search strategies: *term strategies*, *database strategies*, *interaction strategies* and *search strategies*.

Based on empirical studies of library information seekers, Pejtersen (1989) identified three high-level ISSs: *analytical search, search by analogy,* and *browsing*. This classification scheme was employed to design an interactive library system named “Book House”, demonstrating a practical aspect of the scheme.

Chen and Dhar (1991) identified five types of strategies by exploring cognitive processes of users: the *known-item- instantiation strategy,* the *search-option-heuristics strategy,* the *thesaurus-browsing strategy,* the *screen-browsing strategy,* and the *trial-and-error strategy*.

These studies identified the varieties of tactics/moves or ISSs employed by the users during their information-seeking processes. But they are limited at explaining which conditions these ISSs can apply to and how to characterize ISSs in multi-dimensions.

Chang (1995) identified four underlying common dimensions of browsing: *the level of scanning activity, the specificity of information provided by the resource, the definiteness or specificity of the patron’s goal, and the specificity of the object sought.*

Belkin et al. (1993) developed a multi-faceted classification scheme of ISSs to characterize information-seeking behaviors. Their classification scheme had four facets: *method of interaction* (scanning-searching), *goal of interaction* (learning-selecting), *mode of retrieval* (recognition-specification), and *resource considered* (information-meta-information). In this scheme, the method of interaction can be characterized based on the basic distinction between *searching* for a specific known item, or *scannin*g for interesting items from information resources. Searching for identified items can be characterized as *retrieval by specificatio*n, while exploring relevant items can be characterized as *retrieval by recognitio*n. The goal of interaction could be *learning* about different aspects of an information item, or *selecting* interesting items for retrieval. The resource may be interacting with *information* items or with *meta-information* about the structure of information items. Sixteen distinct kinds of ISSs were identified within this multi-dimensional space (see Figure 2.5). Information-seeking behaviors can be characterized by the movement from one ISS to another during the information-seeking process.

ISS	**Method**		**Goal**		**Mode**		**Resource**	
	Scan	**Search**	**Learn**	**Select**	**Recognize**	**Specify**	**Infor-mation**	**Meta-information**
1	X		X		X		X	
2	X		X		X			X
3	X		X			X	X	
4	X		X			X		X
5	X			X	X		X	
6	X			X	X			X
7	X			X		X	X	
8	X			X		X		X
9		X	X		X		X	
10		X	X		X			X
11		X	X			X	X	
12		X	X			X		X
13		X		X	X		X	
14		X		X	X			X
15		X		X		X	X	
16		X		X		X		X

Figure 2.5 Multi-dimensional classification of ISSs (after Belkin et al., 1993)

In Figure 2.5, for example, ISS 2 could be a situation in which a person wants to learn about the characteristics of the information space. The person might look through the meta-information resource (e.g., a thesaurus) to learn about the overall structure of the information space. This ISS is associated with an unformulated information need.

Xie (2000) researched 40 user cases selected from four types of libraries and identified two dimensions of ISSs: *methods* and *resources*. *Methods* are the techniques users employed during the interaction process, such as scanning, searching, tracking, selecting, comparing, acquiring, consulting and trial and error. *Resources* include information, information objects, and human. Xie demonstrated that ISSs could be characterized by multiple combinations of eight types of methods and six types of resources. Based on this, she identified three types of shifts of ISSs: *change of methods, change of sources,* and *change of both methods and resources*.

After investigating the ordinary work of knowledge workers, Cool and Belkin (2002) proposed a multi-faceted classification scheme (see Table 2.2). Five major facets were identified: *communication behaviors, information behaviors, objects interacted with, common dimensions of interaction,* and *interaction criteria*. This scheme describes interactions by combining the elements in different facets.

In summary, the above studies identified the moves, tactics or ISSs users employed during the information-seeking processes (see Table 2.3). Studies on moves or tactics focused more on users' decisions and activities, while studies on ISSs paid more attention to users' motivations and high-level approaches. The move/tactic-related studies tended to be at a theoretical level, while the ISS-related studies were more practical. An example of the latter is BRAQUE, a system established on the basis of the multi-dimensional classification scheme of Belkin et al. (1993).

Table 2.2

Facets of a Classification of Interactions with Information (after Cool & Belkin, 2002)

Facets	Sub-facets	Properties	Values
Communication behaviors	Medium		Speech, text, video, …
	Mode		Face-to-face, mediated, …
	Mapping		One-to-one, one-to-many, many-to-many
Information behaviors	Create		
	Disseminate		
	Organize		
	Preserve		
	Access	Method	Scanning…searching
		Mode	Recognition…specification
	Evaluate		
	Comprehend		Read, listen
	Modify		
	Use		Interpret
Objects interacted with facet	Level		Information, meta - information
	Medium		Image, written text, speech, …
	Quantity		One object, set of objects, database of objects
Common dimensions of interaction	Information object		Part – whole
	Systematicity		Random – systematic
	Degree		Selective – exhaustive
Interaction criteria			Accuracy, alphabet, authority, date, importance, person, time, topic, …

Table 2.3

Categories of Research in ISSs

Category types	Literatures	Examples/*Dimensions*
Moves	Fidel (1985)	Operational moves, and conceptual moves
Tactics	Bates (1979)	Monitoring tactics, file structure tactics, search formulation tactics, and term tactics
	Harter and Roger-Peters (1985)	Overall philosophical attitudes and approach, language of problem description, record and file structure, concept formulation and reformulation, recall and precision, and cost/efficiency
	Shute and Smith (1993)	Delete a slot (that was ANDed) from the current topic add a broader slot-filler to a slot already, represented in the current topic (using OR), add a sibling slot-filler to a slot already, represented in the current topic (using OR), add a cousin slot-filler to a slot already represented in the current topic (using OR), add a slot-filler to a slot that is not filled in the current topic (using OR), add a slot-filler for a slot that is not represented in the current topic (using AND), delete a slot-filler that is represented in the current topic using OR, exclude a slot-filler (using NOT), add a slot-filler for a slot that is already represented in the current topic (using AND),replace a slot-filler with a narrower, slot-filler in the same slot Eliminate a slot-filler from one slot and instead add a slot-filler in a new slot, replace a slot-filler with a sibling slotfiller (in the same slot), and replace a slot-filler with a cousin slotfiller (in the same slot)
ISSs	Belkin et al. (1996)	Term strategies, database strategies, interaction strategies, and search strategies
	Chen and Dhar (1991)	The known-item instantiation strategy, the search-option heuristic strategy, the thesaurus-browsing strategy, the screen-browsing strategy, and the trial-and-error strategy
	Pejtersen (1989)	Analytical search, search by analogy, and browsing
	Marchionini (1995)	Analytical strategies and browsing strategies
	Chang (1995)	*The level of scanning activity, the specificity of information provided by the resource, the definiteness or specificity of the patron's goal, and the specificity of the object sought*
	Belkin et al. (1993)	*Mode of retrieval (recognition-specification), method of interaction (scanning-searching), goal of interaction (learning-selecting), and resource considered (information-meta-information)*
	Xie (2000)	*Methods (scanning, searching, tracking, selecting, comparing, acquiring, consulting and trial and error), and resources (information, information objects, human)*
	Cool & Belkin (2002)	*Communication behaviors, information behaviors, objects interacted with, common dimensions of interaction, and interaction criteria*

2.4 Models of Information-Seeking Dialogue

Research in information-seeking dialogues models the interaction between the user and the system at the discourse act level. Human-Computer interaction can be modeled based on conversational patterns of human-human interaction. These conversational patterns represent different kinds of "semantic and logical relations that can hold between utterances of a discourse" (Reichman, 1985, p. 35).

The "Conversation for Action" (CfA) model (Winograd & Flores, 1986) introduces sequences of *dialogue acts* and decides how they interplay in progressive *dialogue states*. The model is described as the traversal of a state-transition network (see Figure 2.6). In this model, the arcs represent speech acts, and the nodes represent dialogue states. The dialogue starts with partner A's "request", which then could be followed by B's "promise" to comply; B's "counter" to propose a different action; B's "reject" to comply; or A's "withdraw" of the previous request.

The "*Co*nversational *R*oles" (COR) model proposed by Sitter and Stein (1992) is a formal model of information-seeking dialogues. This model is derived from the CfA model by Winograd and Flores (1986) with some changes. Full details of this model are discussed in Chapter 8.

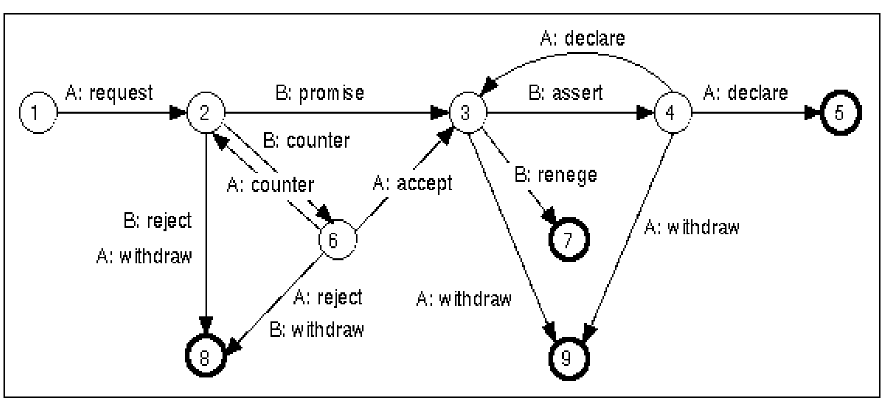


Figure 2.6 The basic "Conversation for Action" (after Winograd and Flores, 1986)

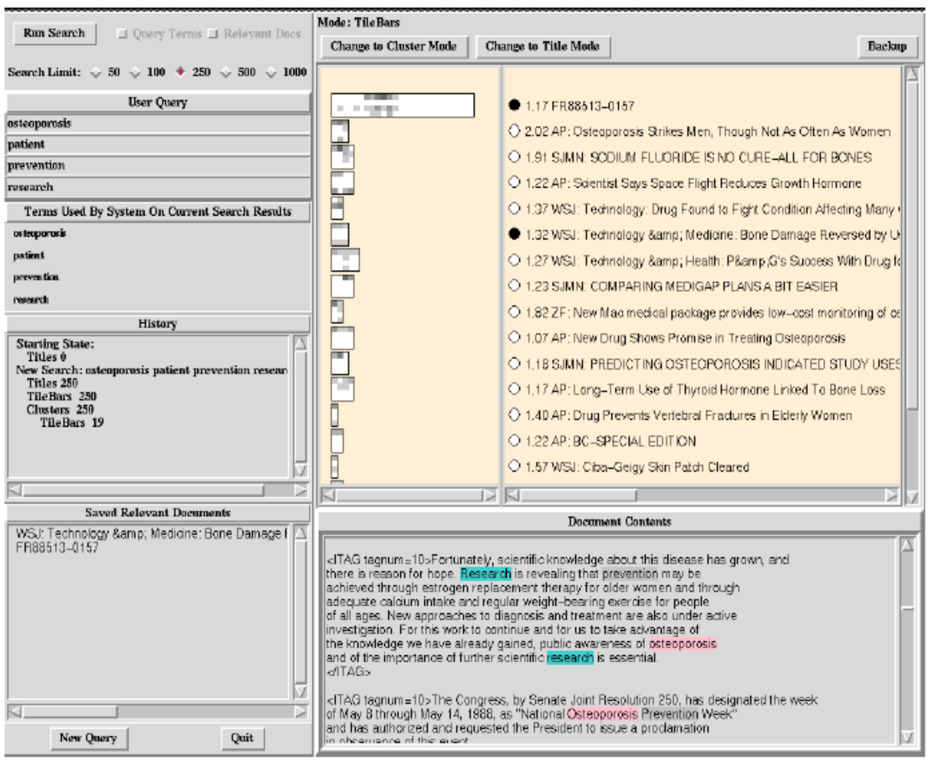
2.5 Integrated IR Systems

It is widely recognized that traditional information systems based solely on searching through specification leave room for improvement in the support of information-seeking in general. Some work has been done to investigate the possibilities of integrating separate IR systems, separate retrieval models, or multiple ISSs into a single system framework.

Frisse and Cousins (1989) developed a system which integrated hypertext and probabilistic retrieval models. The user interface of this system incorporated two general navigational methods in hypertext – local and global. The local navigation method navigates through "document space" while the global method navigates through "index space." Text buttons are used to move through document space, whereas relevance feedback is adopted to move across index space through selection of buttons labeled "Like" and "Don't like." The probabilistic inference techniques are applied to hierarchical index spaces. One problem of this system is that it does not allow users to proceed through index space unless they specify a change in topic.

Hearst et al. (1996) integrated a browsing system *Scatter/gather* (Cutting, Karger, Pedersen & Turkey, 1992), a vector-space best-match retrieval system, and the visualization system *TileBars* (Hearst, 1995) to provide users with effective support for choosing relevant items. The integrated system offers users multiple modes to view the retrieval results. Figure 2.7 displays the interface in TileBars mode, Figure 2.8 shows the interface in title mode, and Figure 2.9 displays the interface in cluster mode. The TileBars mode interface presents the user's query, the system's interpretation of the query, a log of the history, and information about the saved documents, and enables the user to change the mode of the retrieval results display. In cluster mode, the retrieved results are categorized into four clusters labeled with topical terms which indicate the central topics covered in the documents in each cluster.

These systems integrated separate IR models or systems. But completely combining two separate models makes it difficult to optimally support different ISSs, and integrating originally different systems makes it hard to use the results related to one ISS to support another.



<u>Figure 2.7</u> The Interactive Interface, in TileBars mode (after Hearst et al., 1996)

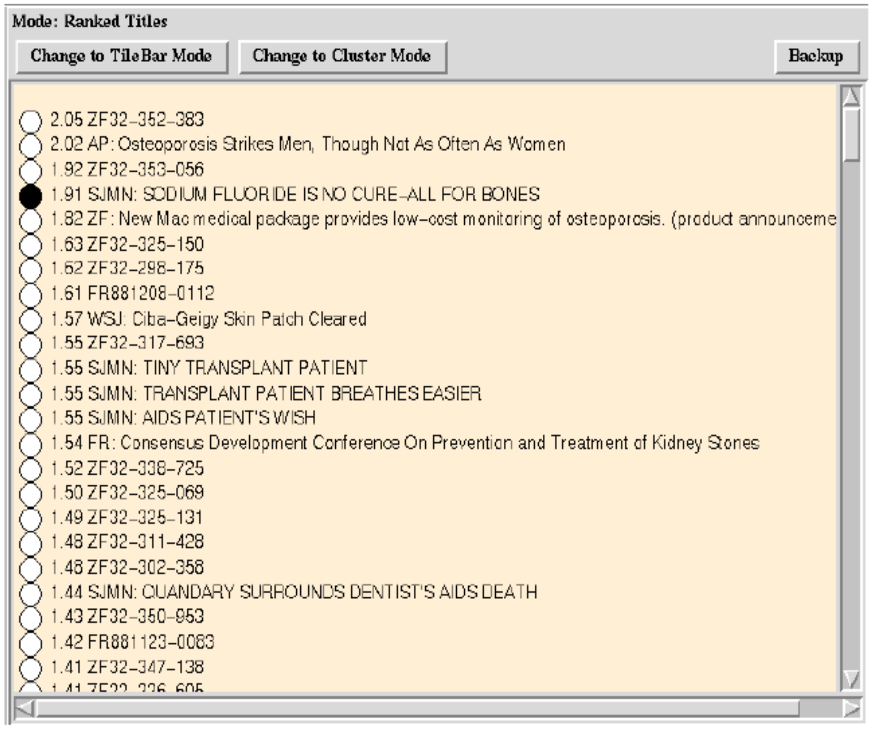


Figure 2.8 The interface in title mode (after Hearst et al., 1996)

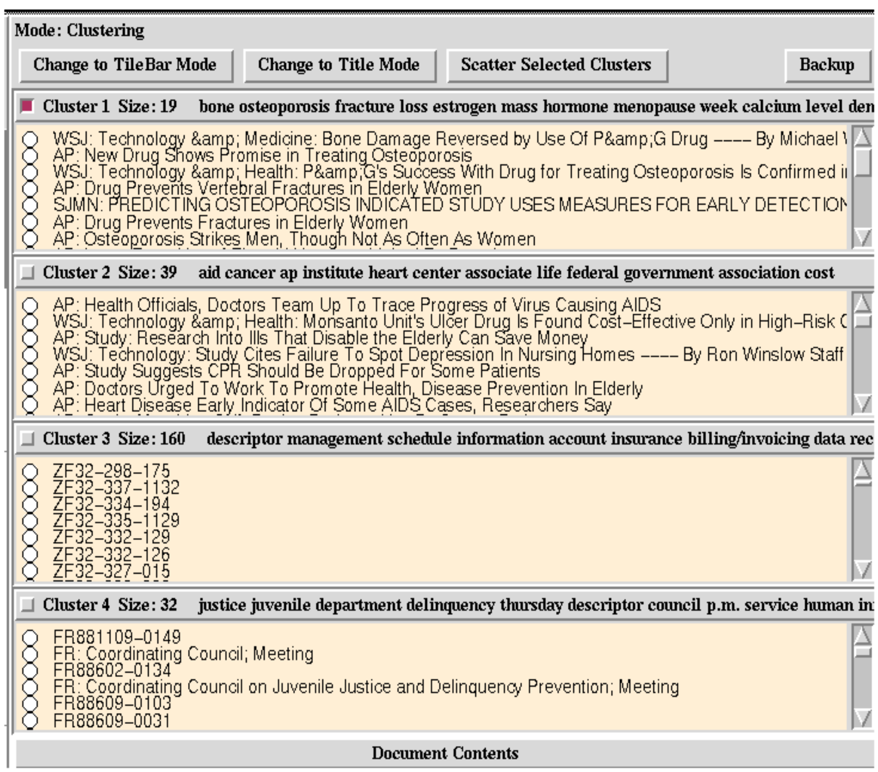


Figure 2.9 The interface in cluster mode (after Hearst et al., 1996)

Thompson and Croft (1987) proposed a system named *I^3R* to support both browsing and specified searching. The I^3R system employed domain knowledge to refine the model of the user's information need and provided a browsing mechanism to enable the user to navigate through the knowledge base. Query formulation and refinement, search, and user evaluation are the three major phases of a typical session in the I^3R system. In the query formulation and refinement phase, the system frames a precise description of the information need, which is referred to as the *request model*. The information in the request model is then used to retrieve documents in the search phase. During the user evaluation phase, the user reviews the retrieved document list and identifies relevant documents. Browsing plays important roles in all the three phases. The browsing expert helps the user navigate through the knowledge base which is graphically displayed as a network of nodes and links.

Belkin et al. (1993) designed an IR system called *BRAQUE* (BRowsing And QUEry formulation) which supported multiple ISSs and enabled seamless movement from one specific ISS to another. In this system, a user can choose either to search or browse (see Figure 2.10). When the user is uncertain about how to formulate a query, or does not have a known document to search for, s/he might type in a query such as "natural language information retrieval" as the starting point for browsing (see Figure 2.11). Because this query is not in the thesaurus, the system displays some thesaurus terms found in the documents retrieved by the query (see Figure 2.12). Among these terms, the user is particularly interested in the term "information retrieval," so s/he continues to browse on this term. Figure 2.13 shows the thesaurus display for the term "information retrieval". The user then chooses to view the documents related to "query languages" (see Figure 2.14). Figure 2.15 shows the full citation of the selected document.

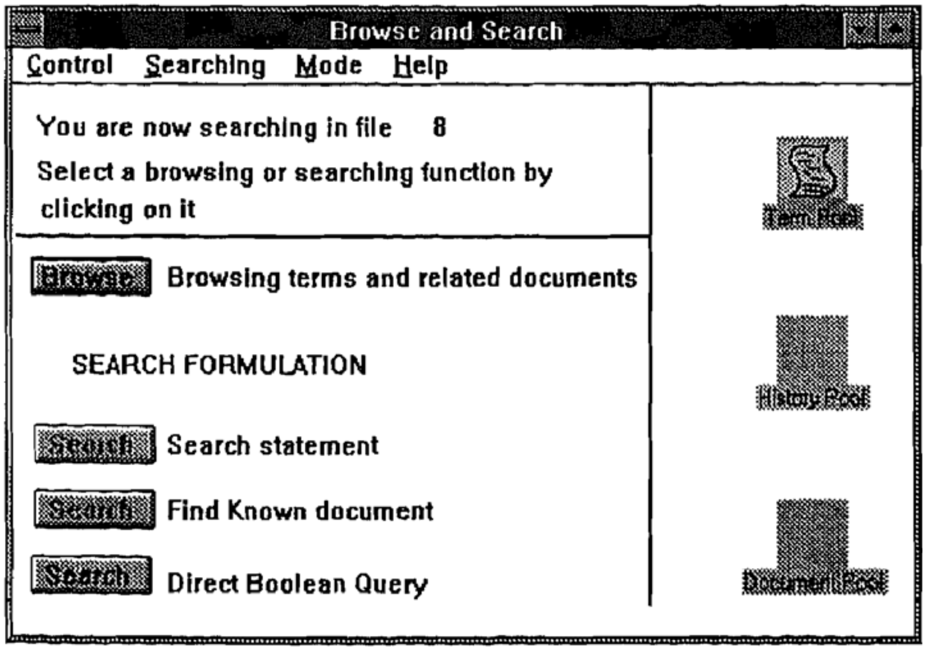


Figure 2.10 Main browse and search menu in BRAQUE (after Belkin et al., 1993)

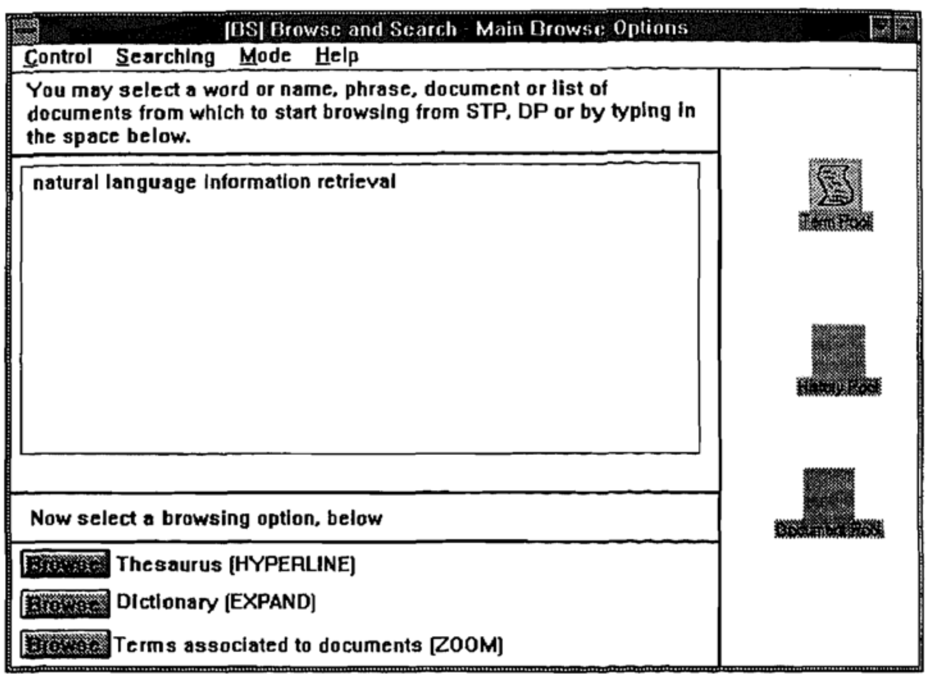


Figure 2.11 Windows for beginning a browse (after Belkin et al., 1993)

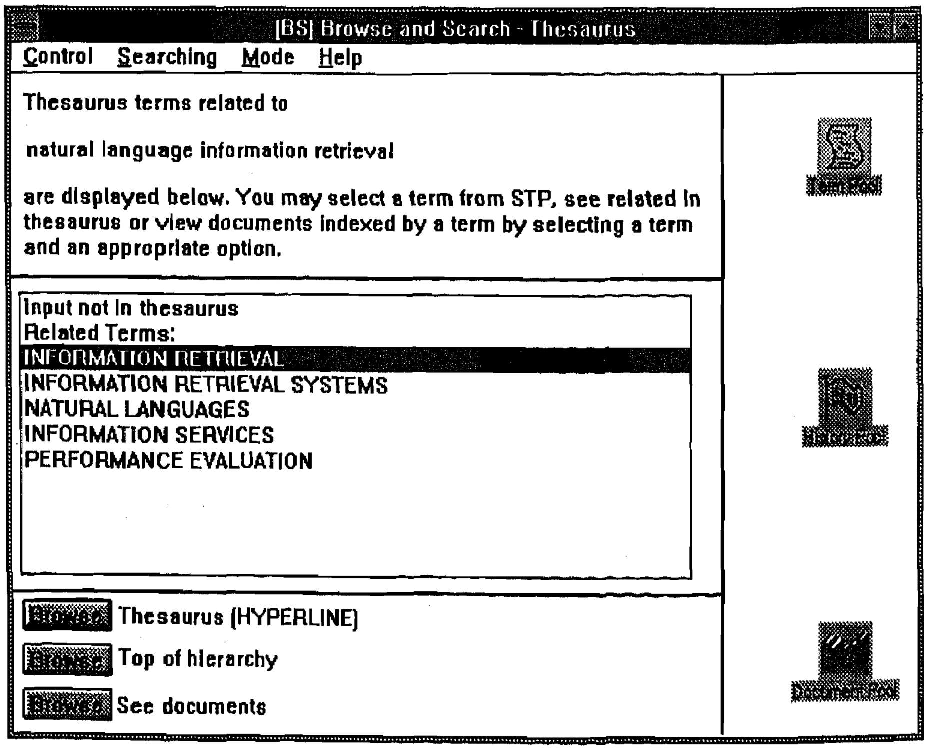


Figure 2.12 Thesaurus entries related to input phrase (after Belkin et al., 1993)

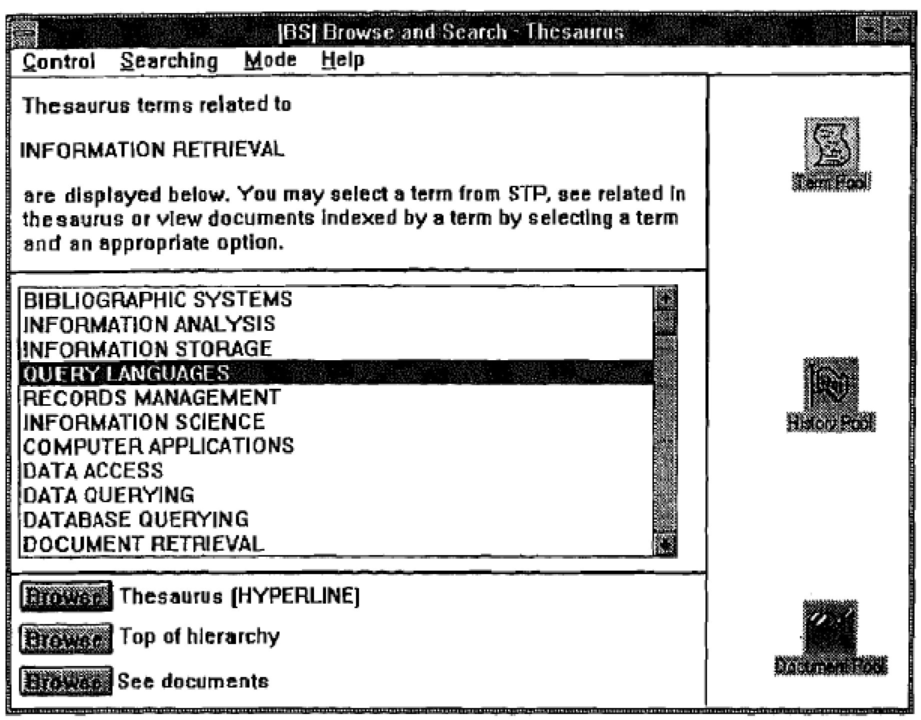


Figure 2.13 Thesaurus display for "Information Retrieval" (after Belkin et al., 1993))

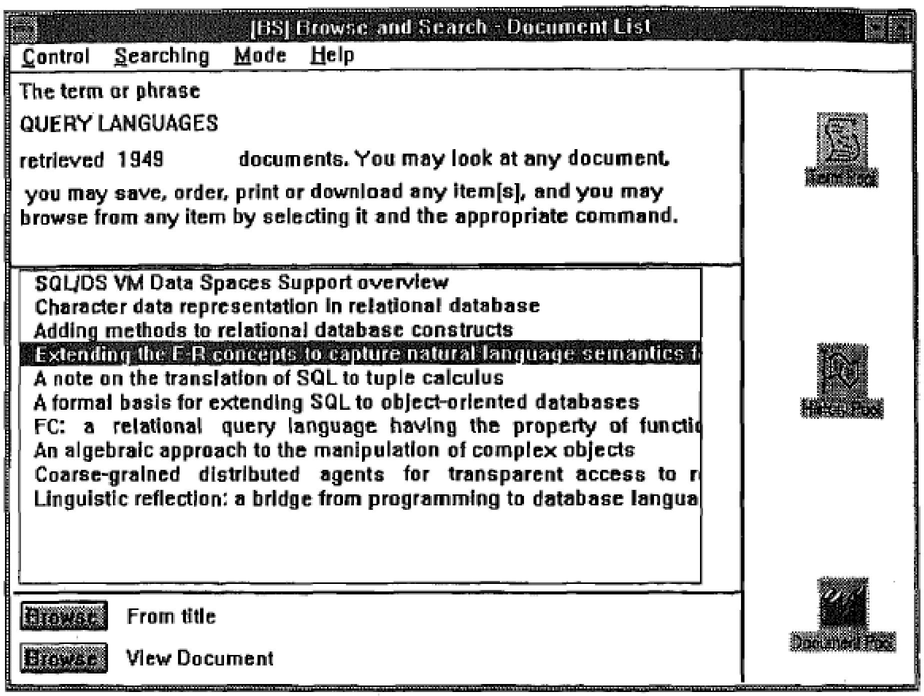


Figure 2.14 Titles of documents indexed by "Query Languages" (after Belkin et al., 1993)

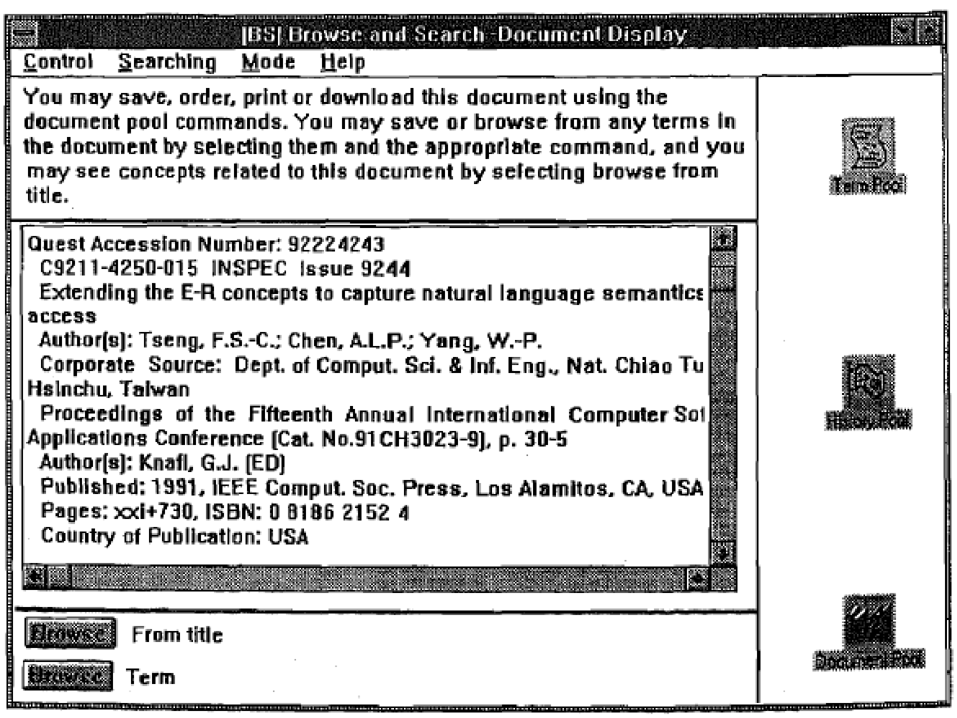


Figure 2.15 Top of single document text display (after Belkin et al., 1993)

Belkin et al. (1995) designed a dialogue-based interactive system, *MERIT*, which supported several ISSs and modeled changes of ISSs using dialogue structures. The basic principle for the MERIT system design is that IR interaction can be regarded as a "conversation" between the user and the system. In a situation that a user wants to find some projects in a particular field, the system displays the retrieved 28 projects in overview in which projects and their funding programs are listed (Figure 2.16). MERIT offers several options for the user to respond: look at one item in detail, look for interrelations, modify the query, or pose a new query. In a situation that the user recognizes one relevant item and wants to select it, the user fills in the term "analogical reasoning" in the query form then clicks "search & show results" button (Figure 2.17). The system then informs the user that no project has been found and suggests some other options to search.

These systems (I^3R, BRAQUE, MERIT) integrated multiple ISSs in a single framework, but their effectiveness remains to be evaluated empirically.

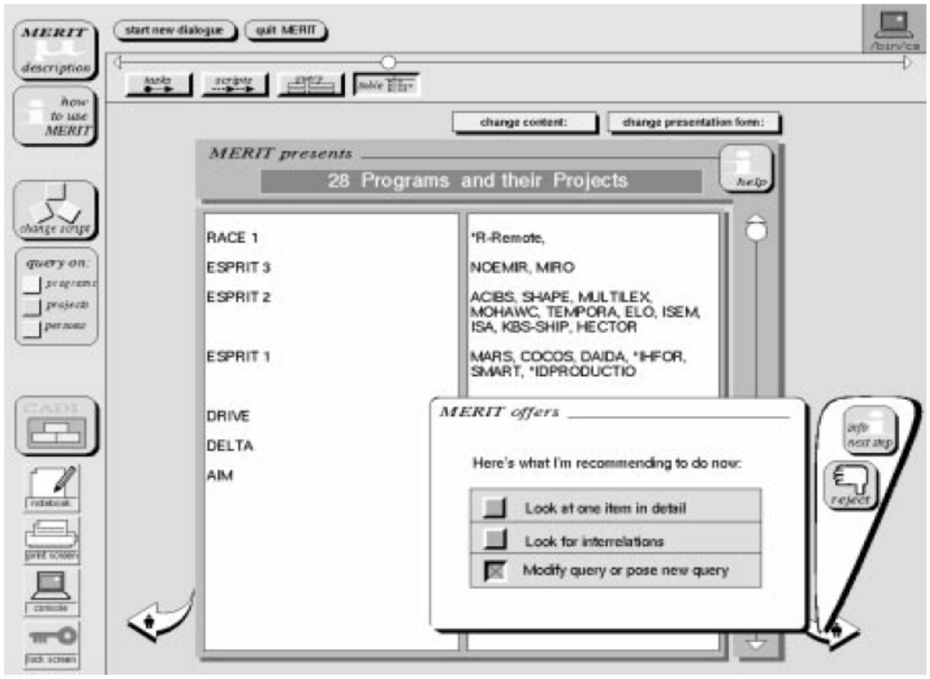


Figure 2.16 Retrieved items and a system's offer (after Belkin et al., 1995)

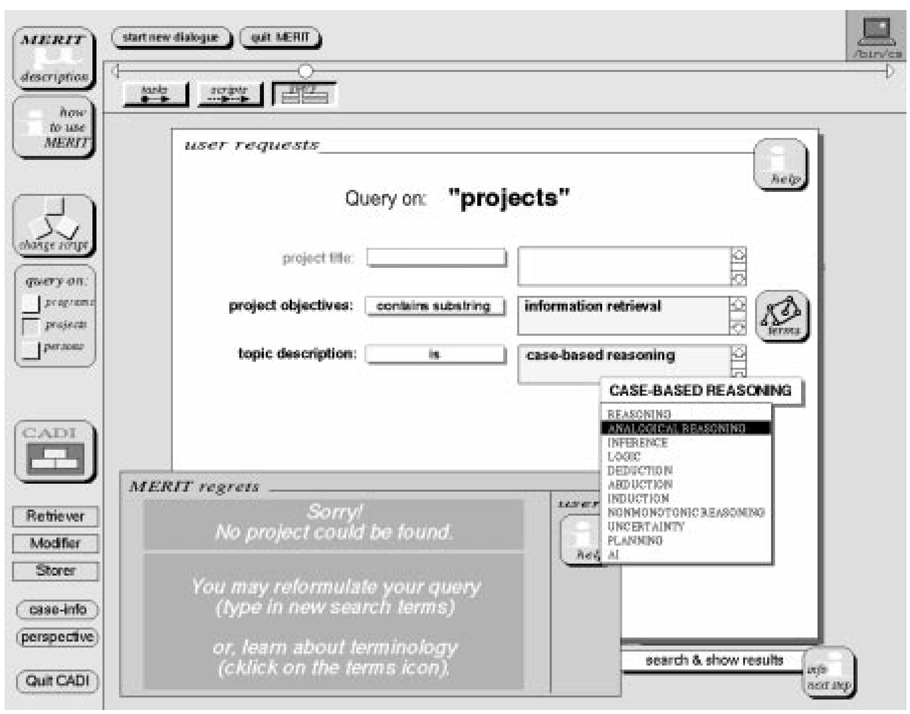


Figure 2.17 Query and display of a ranked list of search terms (after Belkin et al., 1995)

The *ScentTrails* method developed by Olston and Chi (2003) integrated browsing and searching to help people get information on the web. It used hyperlink highlighting to indicate a path to retrieval results so users can use both browsing and searching by employing a combination of browsing cues (e.g., snippets) and the search cues offered by the link highlighting. Information scent is the "imperfect, subjective perception of the value, cost, or access path of information sources obtained from browsing cues" (p. 181).

In this system, users can type a list of search terms into an input box at any time while they are browsing. These search terms represent the user's partial information goal, or the portion of the user's information goal. In ScentTrails, links are highlighted by considering the relevancy to the user's partial information goal, as well as the number of clicks to relevant pages.

In Figure 2.18, the web page highlighted the partial information goal "remote diagnostic technology." Here, link highlighting is created by using the increased font size of the link anchor text. Different sizes of each link anchor text show to what extent the pages related to that link match the partial information goal. For the remaining information goal, that is, finding copiers with the speed of at least 75 copies per minute, the browsing cues and search cues need to be considered at the same time.

The ScentTrails system was evaluated in a preliminary user study using a within-subjects experiment. In this study, ScentTrails was compared with three other interfaces which included one similar interface, one browsing system and one search system. The results showed that ScentTrails enabled people to find information more quickly than by searching or browsing alone. However, due to the small sample size (12 subjects), the generalizability of the results is unclear, and the method remains to be implemented and tested in a larger context.

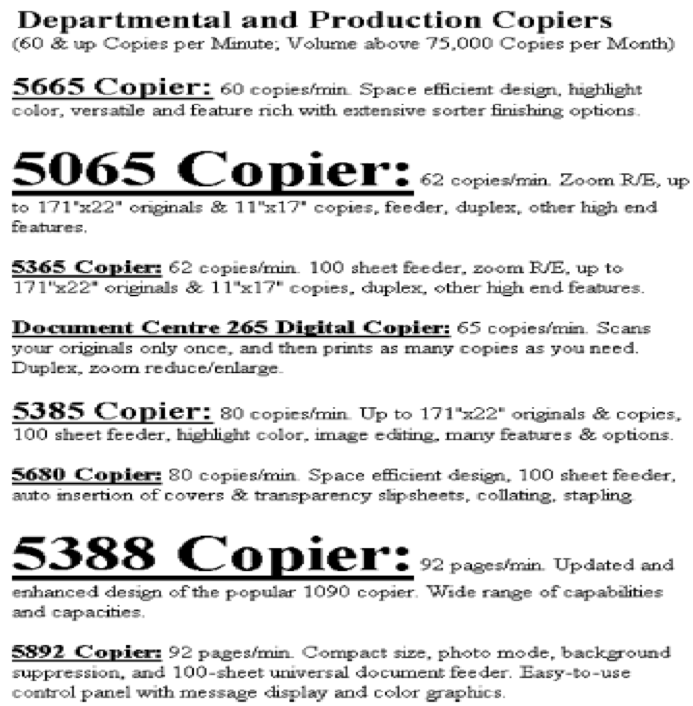


Figure 2.18 A web page whose link anchors have been highlighted for the partial information goal “remote diagnostic technology” (after Olston & Chi, 2003)

The *Phlat* system (Cutrell, Robbins, Dumais, & Sarin, 2006) was designed to facilitate personal information search by integrating keyword search and metadata browsing through different kinds of cues. Phlat also offers a unified tagging mechanism to organize information. The Phlat interface is composed of 3 main visual areas: Query, Filter, and Results. The Query Area shows the current query and the number of search results matching it. The Filter Area comprises 6 buttons for different faceted metadata: saved queries, date, tags, path, People and type. The Results Area shows the results in a columnar list (see Figure 2.19). Two-hundred-twenty-five users tested Phlat for 8 months and the results from usage logs and user feedback showed that Phlat was successful in assisting users in locating personal information. But it is still unknown whether this interface can be successfully used to deal with “non-personal” information.

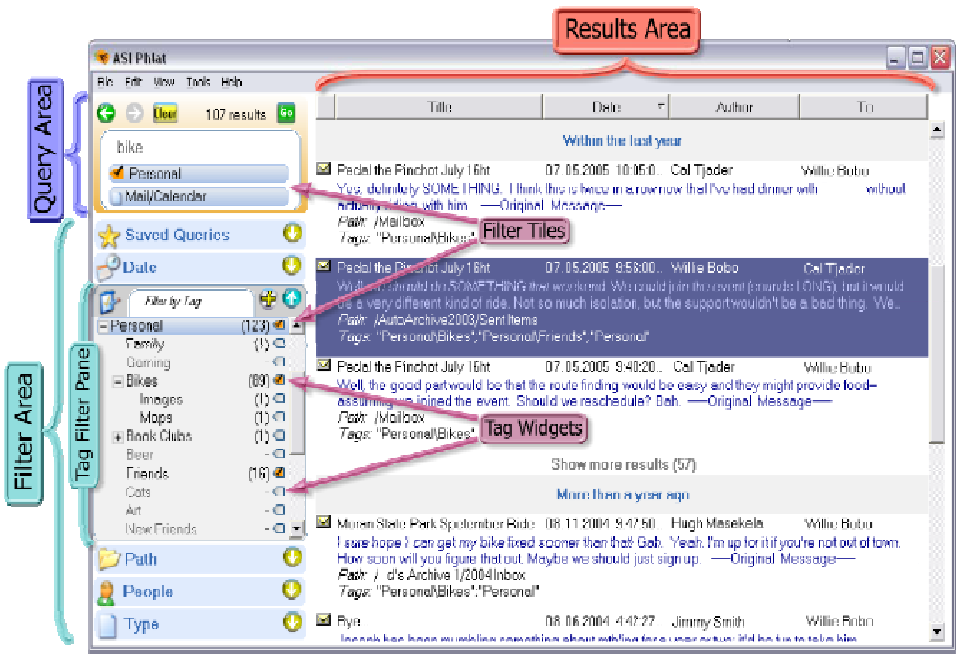


Figure 2.19 The Phlat interface (after Cutrell et al., 2006)

The Relation Browser (Marchionini & Brunk, 2006) was built to identify the relationships of a variety of attribute sets, and to better understand the corpus by allowing investigation of multiple “slices” which are defined by attribute value juxtapositions. It couples searching and browsing by providing visualized category overviews of an information space, while offering filtering and exploration of the result set.

As a hypertext browsing system, the *SuperBook* (Egan et al., 1989) has some basic functions, such as Word Lookup, Table of Contents, and Page of Text. Specifically, the Word Lookup can be used to get all occurrences of any word, word stem or combination of words input by the user. The Table of Contents displays a fisheye-like view of the hierarchical topic headings in the document. The Page of Text displays the text chosen by the user. The first version of SuperBook was evaluated to find factors that affect search difficulty in printed text and SuperBook documents. Ten university students participated in the study and four sets of search questions were given to them. Results showed that SuperBook version 1 is competitive with printed documents but not superior. SuperBook version 2 (see Figure 2.20) was aimed at improving search accuracy and speed. The evaluation of version 2 was conducted in a between-subjects design. Twenty university students participated in the study. Eight questions were assigned. The results showed that SuperBook is superior to printed text.

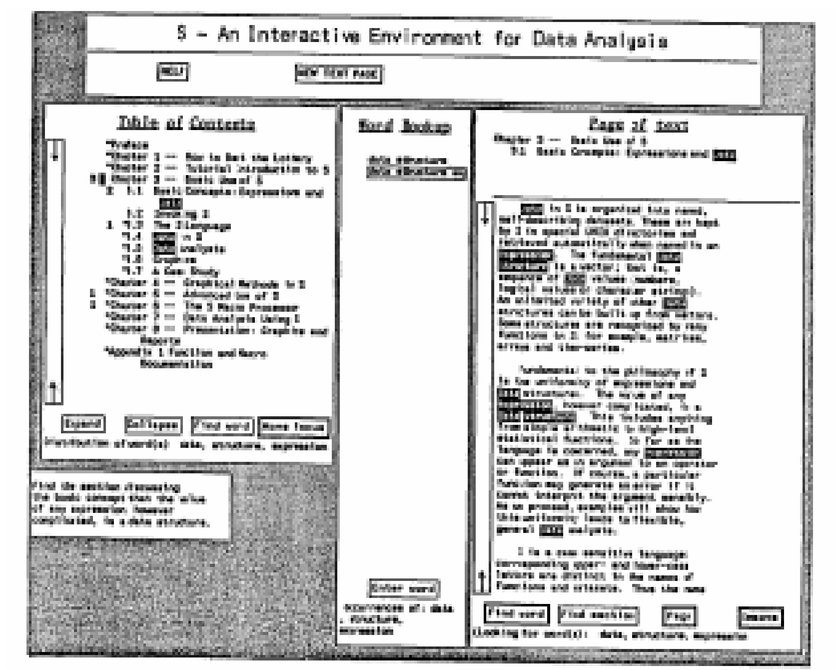


Figure 2.20 Table of contents display (after Egan et al., 1989)

The Flamenco (Flexible Access to Metadata in Novel Combinations) Image Browser (Yee, Swearingen, Li, & Hearst, 2003) employs hierarchical faceted metadata and dynamically generates query previews to enable users to move through large image collections. It categorizes query results into several regions, each of which links a different group of images to a different kind of metadata related to the specified query (see Figure 2.21). Thirty-two subjects joined the usability study to test the Flamenco system, where fine arts collections were used. Results showed that most subjects strongly preferred using Flamenco. These results further indicated that a category-based approach is successful in offering access to image collections.

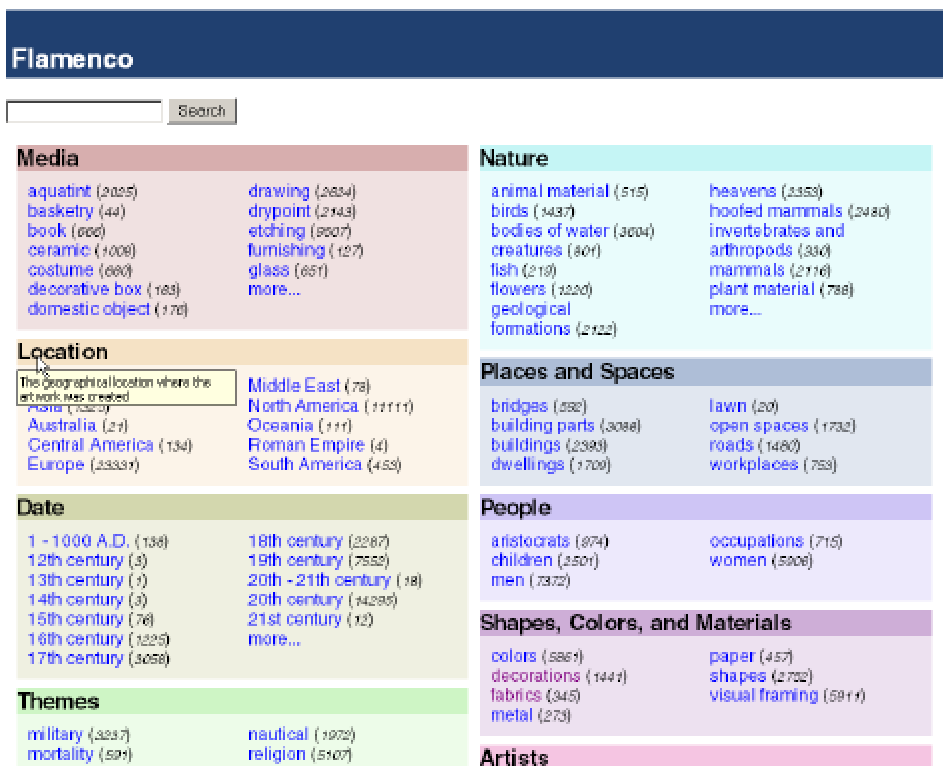


Figure 2.21 Hierarchical faced metadata in Flamenco (after Yee et al., 2003)

Commercial web-based systems such as *Amazon* (http://www.amazon.com) (see Figure 2.22) and *Towers records* (http://www.tower.com) (see Figure 2.23) integrated searching and metadata in order to improve retrieval. For example, besides the search box, Amazon also provides a browse window at the left side which groups the products into a variety of categories. Towers Records categorizes the products into three categories: music, movies and books. It also allows the user to search within categories.

These systems integrated searching with structures of various kinds such as table of contents (SuperBook), category (Relation Browser), and general multi-faceted metadata (Phlat, Flamenco, Amazon, and Towers Records). They served different purposes, such as organizing personal information (Phlat), improving image search (Flamenco), displaying logical relationships of attributes (Relation Browser), and online shopping (Amazon, and Towers Records). The information-seeking approaches users employed (e.g., browsing, searching) were considered in these systems, but multi-dimensions of ISSs were not investigated. Therefore, it is necessary to design a more general IIR system that can support a variety of information-seeking behaviors by characterizing ISSs in multi-dimensions. The effectiveness of such system needs to be empirically tested.

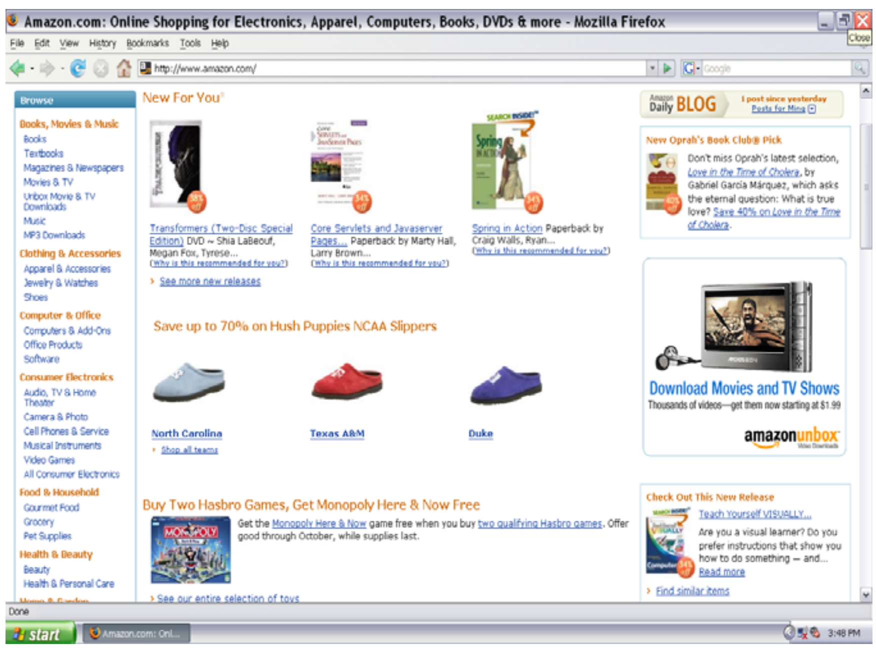


Figure 2.22 Amazon

Figure 2.23 Towers records

Chapter 3

Conceptual Framework

This chapter focuses on the conceptual framework adopted in this study. It starts with the description of the framework, then describes a multi-dimensional classification scheme. Next, it discusses prediction about the optimum combination of IR support techniques. At the end, the definition of the ISSs in this study is introduced.

3.1 Framework

Taking into account the research problems, the conceptual framework for this study:

- Asserts the dominant importance of interaction processes, as opposed to comparison and representation;
- Considers information-seeking context;
- Supports the integration of multiple ISSs in a single system framework;
- Provides a schema for integrating IR systems which can adaptively change from one ISS to another;
- Supports construction of classes of reusable, modular techniques.

The information-seeking episode model proposed by Belkin (1996) (see Figure 3.1) fits many of the above requirements.

In this model, an information-seeking episode is viewed as a sequence of different types of interactions between the user and information objects or IR systems. Each specific interaction is related to contextual factors such as the user's overall situation, current tasks, goals, and intentions. Each of the traditional IR processes (comparison, representation, summarization, navigation, visualization) can be instantiated in a variety of ways (see Table 3.1). Thus in turn each interaction or any particular ISS could be "optimally" supported by different choices of various IR support techniques.

As time passes, a user engages in a variety of interactions which rely on different factors listed above. These interactions include a number of processes such as judgment, interpretation, and modification.

This model suggests that different combinations of IR support techniques could optimally support a specific type of ISS, and an IR system should be able to optimally support different types of interactions or multiple ISSs. This model is the basis for the implementation of IR support techniques for different ISSs in a single system framework.

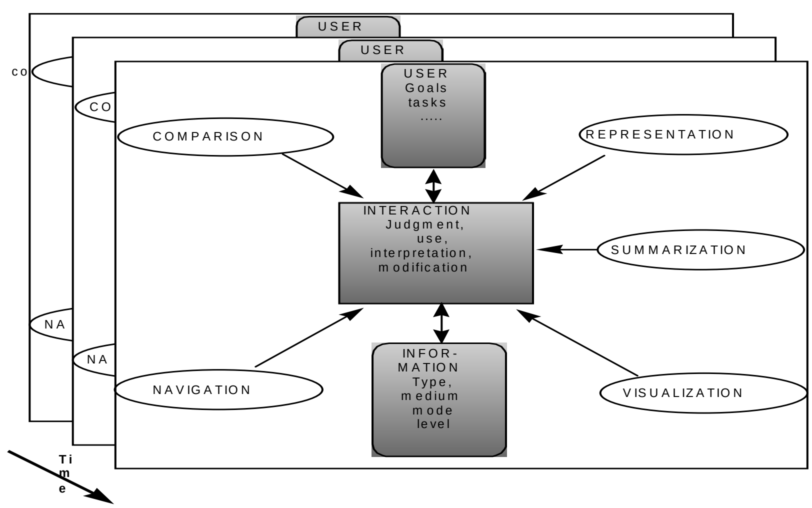


Figure 3.1 Information episode model (after Belkin, 1996)

Table 3.1

Possible IR Support Techniques for Each IR Process

IR processes	Possible IR support techniques
Comparison	Vector space; Exact match
Representation	Clustering; Indexing
Summarization	Titles; Summaries
Visualization	A hierarchy of interconnected structured objects; Lists
Navigation	Following links; Scrolling

3.2 Multi-dimensional Classification of ISSs

Belkin et al. (1993) suggested that the variety of ISSs could be identified by a classification scheme consisting of four binary-valued facets (see Table 3.2). Their claim was that a given ISS could be characterized by a specific combination of values of the four facets in what they characterized as a "space of ISSs" (see Figure 2.5). Any particular ISS then can be described in terms of the combination of its respective values. On the basis of this scheme, the idea of a searcher moving from one ISS to another in this space, and a two-level hypertext IR model, they proposed a design for an IIR system ("BRAQUE") which could support both browsing and querying. Belkin et al. (1995) also employed this scheme to design the "MERIT" system which could support varieties of ISSs with specific dialogue structures. These studies showed that this classification scheme was limited and that more empirical support was needed to evaluate the effectiveness of the systems built upon this model.

The multi-dimensional classification of ISSs developed by Cool and Belkin (2002) is based on the classification scheme proposed by Belkin et al. (1993). This classification scheme has both theoretical and practical appeal, so it was chosen as the basic classification scheme of ISSs in this study.

This scheme (see Figure 3.2) was based on an empirical study of knowledge workers in their regular work environments. It supports interaction between the user and information objects by combining the elements in different facets. These elements could help identify appropriate IR support techniques for each particular ISS. Figure 3.2 displays the relevant facets, and their values, which are the basis for the description of ISSs used in this study. In particular, two basic classes of ISSs were identified: those characterized by the method of *searching*, and those characterized by the method of *scanning*, within the specific information behaviors of *access*.

Information Behaviors Facet (This facet includes a variety of different types of such behaviors; for this study reported here, only the "Access" behavior was considered.)

Access

- Method: Scanning … Searching
- Mode: Recognition … Specification

Objects Interacted with Facet

- Level: Information … Meta-information
- Medium: Image, written text, speech, …
- Quantity: One object, set of objects, database

Common Dimensions of Interaction Facet

- Information object: Part … Whole
- Systematicity: Random … Systematic
- Degree: Selective … Exhaustive

Interaction Criteria Facet

- e.g., accuracy, alphabet, authority, date, person, …

Figure 3.2 A faceted classification of ISSs (after Cool & Belkin, 2002)

Table 3.2

Facets of ISSs (after Belkin et al., 1993)

Facet	Values
Method of interaction	(Scanning; Searching)
Goal of interaction	(Learning; Selecting)
Mode of retrieval	(Recognition; Specification)
Resource interacted with	(Information; Meta-information)

3.3 Scanning vs. Searching

Browsing has been defined in a variety of ways such as “scanning a resource” (Belkin et al., 1993, p. 331); “the process of exposing oneself to a resource space by scanning its content (objects or representations) and/or structure, possibly resulting in awareness of unexpected or new content or paths in that resource space” (Chang, 1993, p. 258); “an approach to information seeking that is informal and opportunistic and depends heavily on the information environment” (Marchionini, 1995, p. 100). These descriptions indicate that the activity of *scanning* is the basic strategy used in the browsing process.

According to Chang (1995), scanning is composed of four levels, that is, looking, identifying, selecting and examining. The act of looking refers to looking through information resources, identifying refers to recognizing interesting items, selecting refers to choosing an interesting item, and examining refers to viewing parts of an item to achieve the specific goal.

Marchionini (1995) defined scanning as “a perceptual recognition activity that compares sets of well-defined objects with an object that is clearly represented in the information seeker’s mind” (p. 111).

This study adopts the classification scheme provided by Cool and Belkin (2002). In terms of this scheme, scanning was associated with selection by recognition, while searching was associated with selection by specification. Table 3.3 lists the facets, sub-facets, properties and values for both scanning and searching strategies.

In order to support these ISSs, IR processes such as comparison, representation, visualization, summarization and navigation play different roles, and take different forms during the information-seeking process. Comparison is a process of finding and ranking items with respect to the person's information problem. The corresponding techniques include exact match, best match (vector space, probabilistic, language modeling), and so on. Representation is about the way that the database contents and information problems are represented. The techniques for representing the contents of the database include automatic indexing, manual indexing, and clustering. Visualization provides descriptions of overall retrieval results. Possible IR support techniques for visualization include ranked lists and graphical depiction of clusters (including interconnected or hierarchical). Summarization provides condensed representations of documents such as titles, abstracts, or other formats. Navigation is the process enabling the person to explore the databases or documents in order to compare them with respect to the information problem. Navigation techniques include following links, scrolling up and down.

Table 3.3

Multi-dimensional Classification of Scanning and Searching

ISSs	Facets	Sub-facets	Properties	Values
Scanning	Information behaviors	Access	Method	Scanning
			Mode	Recognition
	Objects interacted with	Level		Information, meta-information
		Medium		Written text
		Quantity		One object, set of objects, database of objects
	Common dimensions of interaction	Information object		Part - whole
		Systema-ticity		Random - systematic
		Degree		Selective - exhaustive
	Interaction criteria			Accuracy, alphabet, authority, date, …
Searching	Information behaviors	Access	Method	Searching
			Mode	Specification
	Objects interacted with facet	Level		Information, meta-information
		Medium		Written text
		Quantity		One object, set of objects, database of objects
	Common dimensions of interaction	Information object		Part - whole
		Systema-ticity		Random - systematic
		Degree		Selective - exhaustive
	Interaction criteria			Accuracy, topic, alphabet, authority, date, …

3.4 Predictions about the Optimum Combination of IR Support Techniques

There is much evidence that particular combinations of specific IR techniques are more appropriate for supporting some ISSs than others. However, it is still not clear which combinations are most appropriate for which specific ISS. According to the classification scheme of Cool and Belkin (2002), some ISSs can be identified based on the subfacet *Access* of the facet *Information Behaviors*, in combination with three other facets, as indicated in Figure 3.2. Predictions about the optimal combination of IR support techniques for each specific ISS are described in Table 3.4.

3.5 LEMUR Toolkit

This study required a system in which different IR support techniques can be instantiated for different IR processes. For this purpose, the LEMUR toolkit (http://www.lemurproject.org/), an object-oriented framework which can represent multiple IR processes, was used.

LEMUR was designed and implemented for the purpose of facilitating research in language modeling and IR by a joint project of the Center for Intelligent Information Retrieval at the University of Massachusetts and the School of Computer Science at Carnegie Mellon University. LEMUR supports indexing of large-scale text databases, the construction of retrieval systems on the basis of language models and retrieval models, and the development of simple language models for documents, queries, or sub-collections. The underlying architecture of LEMUR was built to support ad hoc and distributed retrieval with structured queries, cross-language IR, summarization, filtering, and categorization. The fact that it allows choice among a variety of indexing and retrieval techniques makes it suitable for the purposes of this study.

Table 3.4

<u>Examples of ISSs and the Corresponding Combination of IR Support Techniques</u>

ISSs	Facets	Sub-facets	Proper-ties	Values	Examples	Possible combi-nations of IR support techniques
Scanning	Information behaviors	Access	Method	Scanning	Learning about the structure of the databases of a system before an information search starts	Summary of databases; scrolling
			Mode	Recognition		
	Objects interacted with	Level		Information, Meta-information		
		Medium		Written text		
		Quantity		One object, set of objects, database of objects		
	Common dimensions of interaction	Information object		Part - whole		
		Systematicity		Random - systematic		
		Degree		Selective - exhaustive		
	Interaction criteria			Accuracy, alphabet, authority, date, …	Finding some comments from a known book	Table of contents navigation; scrolling
Searching	Information behaviors	Access	Method	Searching	Selecting relevant documents from a specified database	Indexing; best match; titles of documents; following links
			Mode	Specification		
	Objects interacted with	Level		Information, meta-information		
		Medium		Written text		
		Quantity		One object, set of objects, database of objects		
	Common dimensions of interaction	Information object		Part - whole		
		Systematicity		Random - systematic		
		Degree		Selective - exhaustive		
	Interaction criteria			Accuracy, topic, alphabet, authority, date, …	Identifying an electronic book from a specified database	Indexing; best match; complete citation of a book; following links

3.6 Extended Information Interaction Model

By considering the information-seeking model by Belkin (1996) and the multi-dimensional classification scheme by Cool and Belkin (2002), as well as the IR support techniques investigated in this study, an extended model (see Figure 3.3) was created. Please note this model is a reflection of one information-seeking episode of Figure 3.1. In this information-seeking episode, the user interacts with the information objects. Each specific kind of interaction is related to current tasks, goals, and ISSs. Each of the IR processes (comparison, representation, summarization, navigation, visualization) can be instantiated in a variety of ways. For example, comparison was instantiated as best match and exact match, representation as indexing, clustering, and fielded query, summarization as complete citation and title, navigation as following links and scrolling, and visualization as table of contents navigation, clustered retrieval results, and database summary. In fielded query, the retrieval results were represented as groups categorized by the different combinations of Boolean search. A user engages in a variety of interactions which include a number of processes such as judgment, interpretation, and modification.

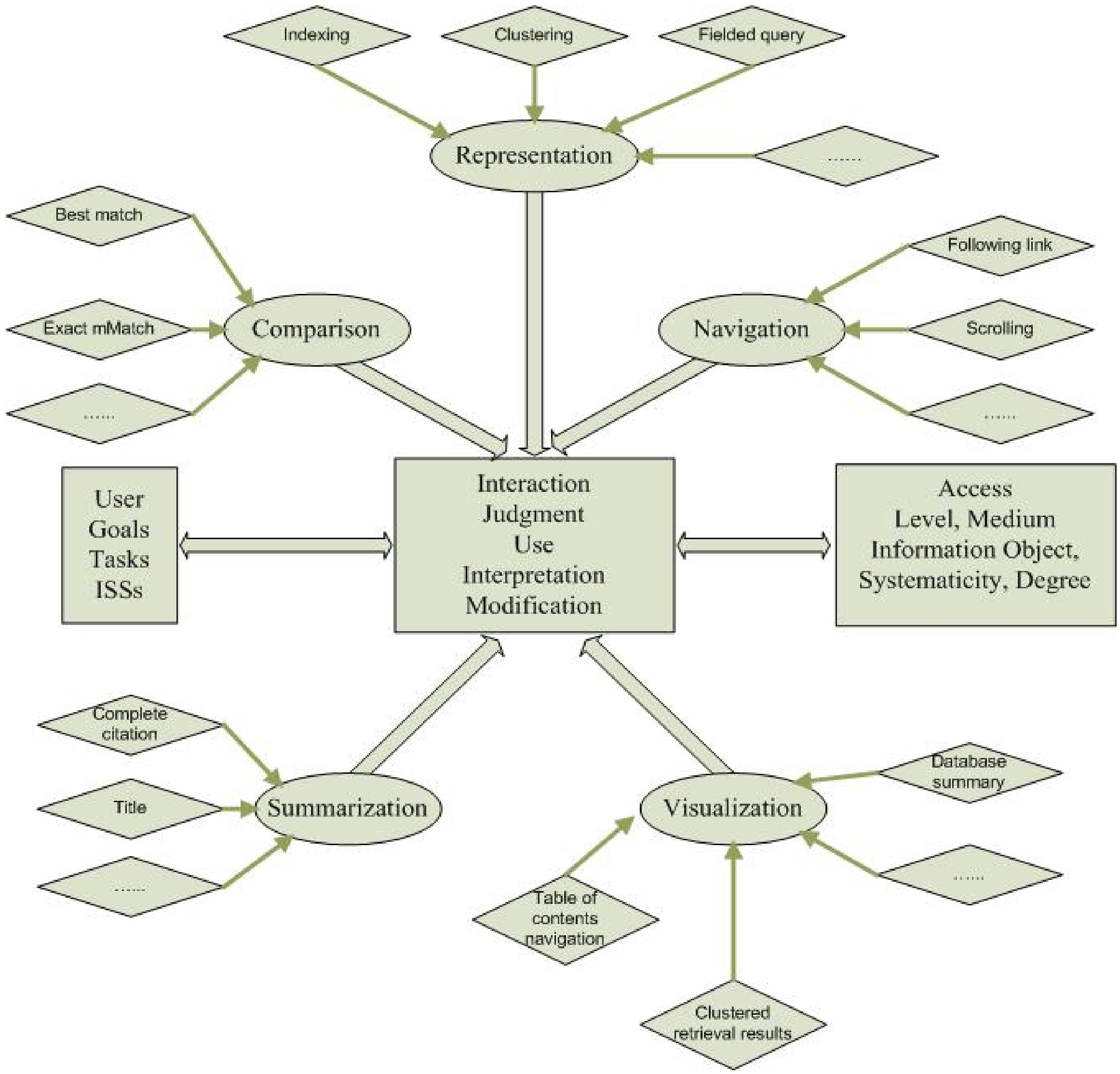


Figure 3.3 Extended information interaction model

Chapter 4

Research Problem 1: Research Method

This chapter describes the research method used in research problem 1. It starts with an overall description of the design of tasks and IR support techniques, then follows with a brief description of system implementation. Hypotheses, tasks and systems are then discussed in more detail. Finally, the experimental design, sampling, measures, data collection, and procedure are introduced.

4.1 Overall Description

Research problem 1 investigated how different IR support techniques affected the performance of different systems under different situations and tasks. The experimental systems were designed by tailoring to several different IR support techniques. The respective baseline systems were designed by following the current standard model of specific query input, and a ranked list of search results.

Two sets of tasks related to document retrieval or book retrieval were devised, each targeted to its own appropriate collection. These two collections are the TREC HARD 2004 collection (Allan, 2005), which is composed of news articles, and a database of electronic books downloaded from Project Gutenberg (http://www.gutenberg.org/).

These tasks are based on and extracted from tasks identified by a cognitive task analysis of IR (Belkin et al., 1993). The two sets of tasks are tailored to the two collections respectively. In each set, there are two tasks representing two situations respectively. These two tasks will be combined into one task in research problem 3.

The cognitive task analysis by Belkin et al. (1993) identified a variety of tasks such as meta-information ("interaction with resources that describes structures and contents of information objects and resources", p. 330), database selection, initial formulation of search topic, query formulation, search strategy formulation, learning ("expanding knowledge of one's goal and problem, the system and resources, the topic" p. 330). In this study, four tasks which are combinations of the tasks identified above were designed. Table 4.1 summarizes the tasks, problems and possible IR support techniques.

In this study, some IR support techniques related to each of the different tasks were investigated (see Table 4.2). Four experimental systems were implemented by considering these IR support techniques, and four different baseline systems were constructed accordingly.

Table 4.1

Tasks, Problems, and Possible IR Support Techniques

Situations	Tasks	Possible related tasks (after Belkin et al., 1993)	Possible problems (after Belkin et al., 1993)	Possible IR support techniques (after Belkin et al., 1993)
Scanning	1. Identify best databases	Database selection	Whether to choose one or several databases	User specification of databases
		Meta-information	Establishing relationships between meta-information and information	Direct manipulation browsing in displays of relationships
		Learning	Knowing about resource contents and organization	User interaction with database description and contents
		Recognition	Getting to the right location in the resource	Display related terms and relationships within database
	2. Find comments from an electronic book	Learning	Identifying appropriate resources	Display of resources available to user, direct user choice based on content description
		Recognition	Getting to the right location in the resource	Display related terms and relationships within database
Searching	1. Find relevant documents	Query formulation	Matching of topic description to effective search statement	Progressive and interactive use of search topic description for query formulation
		Evaluation and reformulation	Relating output to characteristics of search formulation	Ranked document output; Manipulable display of output related to query and search formulation
	2. Find the name of an electronic book	Search strategy formulation	Relating search logic to topic requirements	Provide patterns for search formulation; Structured representation of query and search
		Evaluation and reformulation	Relating output to characteristics of search formulation	Manipulable display of output related to query and search formulation

Table 4.2

The Relations among Situations, Tasks and Systems

ISSs (situations)	Tasks	Experimental systems	Baseline systems	Interested IR support techniques	Corresponding IR support techniques (after Belkin et al., 1993)
1. Scanning	1. Identify best databases	E1.1 (see Figure 4.1) Alphabetically ordered databases with summary for each	B1.1 (see Figure 4.2) Ranked documents with description about which database it is in	Summary of each database	Display related terms and relationships within database
	2. Find comments from an electronic book	E1.2 (see Figure 4.3) Table of Contents navigation within a book	B1.2 (see Figure 4.4) Ranked paragraphs	Table of contents navigation	Display of resources available to user, direct user choice based on content description
2. Searching	1. Find relevant documents	E2.1 (see Figure 4.5) Ranked clusters	E2.1 (see Figure 4.6) Ranked documents	Clustered retrieval results	Manipulable display of output related to query and search formulation
	2. Find the name of an electronic book	E2.2 (see Figure 4.7) Field search	B2.2 (see Figure 4.8) Generic query search	Fielded query	Structured representation of query and search

4.2 Implementing and Evaluating Different Systems for Supporting Specific ISSs

All systems were implemented using the LEMUR Toolkit. As stated previously, LEMUR is designed to support research in language modeling and IR. The toolkit supports large-scale text database indexing, retrieval with structured queries, summarization, filtering and categorization. It is particularly useful in that it provides a flexible platform for researchers to develop their own customizations and applications, which fits very well in this study. In this study, all systems mainly used the Indri retrieval system (Strohman, Metzler, Turtle, & Croft, 2005) and employed particularly the following features: structured-query retrieval, document clustering and passage indexing.

The Indri retrieval model uses both the language modeling and inference network approaches to IR. It can evaluate structured queries using language modeling estimates within the network, rather than *tf.idf* (term frequency-inverse document frequency) estimates. The documents are ranked according to $P(I|D, \alpha, \beta)$, assuming that the information need I is met given document D and hyper-parameters α and β as evidence. The Indri indexing system builds compressed inverted lists for each item and field. The index is self-contained, with all the necessary information to perform queries on that data. When a query is submitted into the Indri system, it is parsed into an intermediate query representation and then passed through a variety of query transformations. The query is evaluated in the following way: first gather the statistics about the number of times terms and phrases appear in a collection, and then use the statistics to evaluate the query against the collection.

For document clustering, the clusters are generated using LEMUR's cluster algorithm, which iterates over the documents in the index and assigns each document to a cluster. It uses centroid-type clusters, with cosine similarity (COS) as the similarity metric, and a minimum similarity score of 0.25 to add a document to an existing cluster. The labels for each cluster are generated using a headline summarization tool developed by Liang Zhou from Information Sciences Institute, University of Southern California, which selects headline words throughout the entire text, and then composes them by finding phrase clusters locally in the beginning of the text. These clusters are ranked based on the posting frequency of the query terms.

Passage indexing is used for searching for specific paragraphs within a book. Each paragraph of each book is treated as a separate document, and the index is generated using the Indri indexing system based on all the paragraphs from all the books. The retrieved paragraphs are limited to at most 500, for ease of display.

In this study, eight systems were designed to support two different ISSs based on the predictions discussed above, with four experimental systems and four corresponding baseline systems. Each experimental system was designed by tailoring to one specific IR support technique, while each baseline system was designed to follow the current standard model of specific query input, and a ranked list of search results. The effectiveness of each experimental system was evaluated by conducting a controlled experiment in comparison with the relevant baseline system. Four experiments were conducted.

4.3 Hypotheses

Hypothesis 1: A system summarizing each database performs better in supporting scanning tasks than a baseline system providing ranked lists of documents with descriptions about which databases these documents are in. (E1.1/B1.1)

Hypothesis 2: A system with table of contents navigation performs better in supporting scanning tasks than a baseline system with a list of ranked paragraphs. (E1.2/B1.2)

Hypothesis 3: A system with clustered retrieval results performs better in supporting searching tasks than a baseline system with a ranked list of retrieval results. (E2.1/B2.1)

Hypothesis 4: A system with fielded query performs better in supporting searching tasks than a baseline system with a generic query search. (E2.2/B2.2)

4.4 Situations and Tasks

4.4.1 Situation 1 (Scanning), Task 1 (T1.1, Identify best databases)

T1.1: A person is interested in one particular topic but has no idea about which of many possible databases to search.

Description: Given this situation, this person needs to identify the best databases for the topic; that is, rank them. To accomplish this, s/he needs to use scanning, as explained below, so the system needs to provide IR support techniques for scanning. This person can then compare the descriptions of the contents of different databases in order to choose the appropriate ones.

Since the person does not know which databases are good, s/he needs to *scan* the *meta-information* about the databases in order to *recognize* the best databases for the topic of interest. In order to get some meta-information of the databases, this person issues a query. That query could be compared using a best match technique against the index terms associated with each database. Based on the meta-information, the person chooses the best databases. To this end, it might be helpful to display the meta-information in such a way that the person can easily discover to which extent the query topic has been covered. A good way to accomplish this is to represent the database by the posting frequency of the index terms. Then the person can see how many documents in the databases are possibly related to the topic of interest by virtue of containing the query terms. Each database is summarized by the number of documents indexed by the terms, and by some description of the contents based on most frequent indexed terms.

This representation will allow the person to compare the different databases and decide which ones look more interesting by scrolling through them.

From the theoretical framework (see Figure 3.1), information-seeking behavior can be seen as the movement from one ISS to another. Different combinations of IR support techniques could optimally support a given type of ISS. In this task, a combination of IR support techniques such as best match, database summary, indexing and scrolling are used to support a scanning strategy. The interactions between the user and the system are related to the situation, task and goal.

4.4.1.1 System Design (E1.1/B1.1)

According to the description above, the experimental system (see Figure 4.1) and baseline system (see Figure 4.2) were designed as follows.

The aim is to compare whether it makes a real difference in performance if descriptions of databases are provided. The experimental system provides the user a list of descriptions about the databases, while the baseline system gives a ranked list of documents retrieved with respect to the topic, from a set of databases, with notation about which database each document is in.

The experimental system lists the descriptions of eight databases from the HARD 2004 collection. For each database, it shows the total number of related documents within that database (limited to at most 100), and the number of related documents for each keyword. Users can click on the database link, which will direct them to another screen with a ranked list of all the related documents in that database. Users can view the complete document by clicking on the link of that document. The baseline system simply provides a ranked list of all the related documents from all the eight databases. For each document, it is noted which database the document comes from. Again, users can view the complete document by clicking on its link.

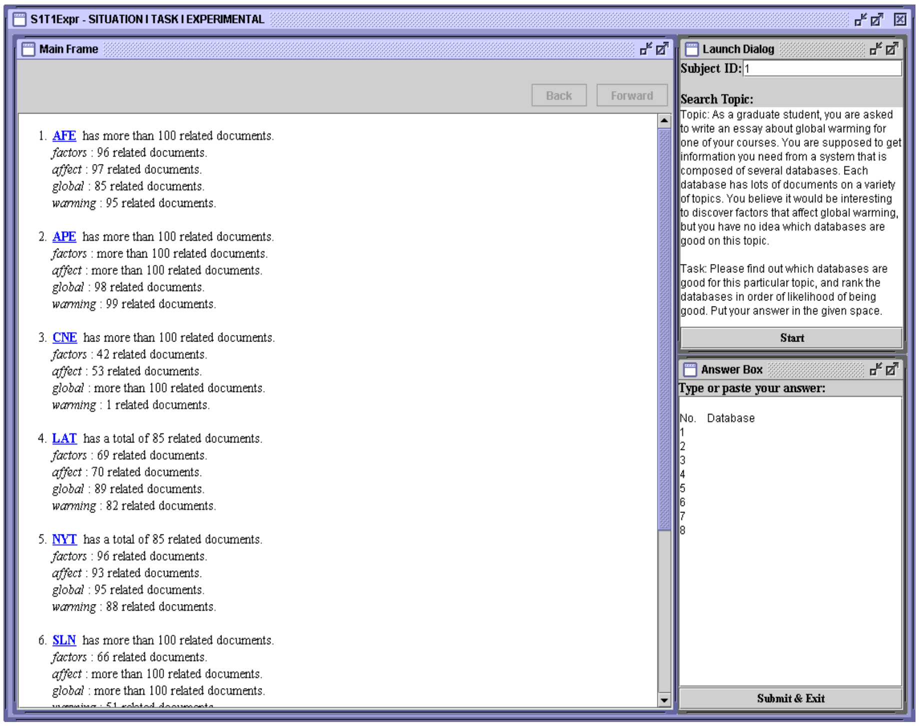


Figure 4.1 Experimental system E1.1 (Situation 1, Task 1)

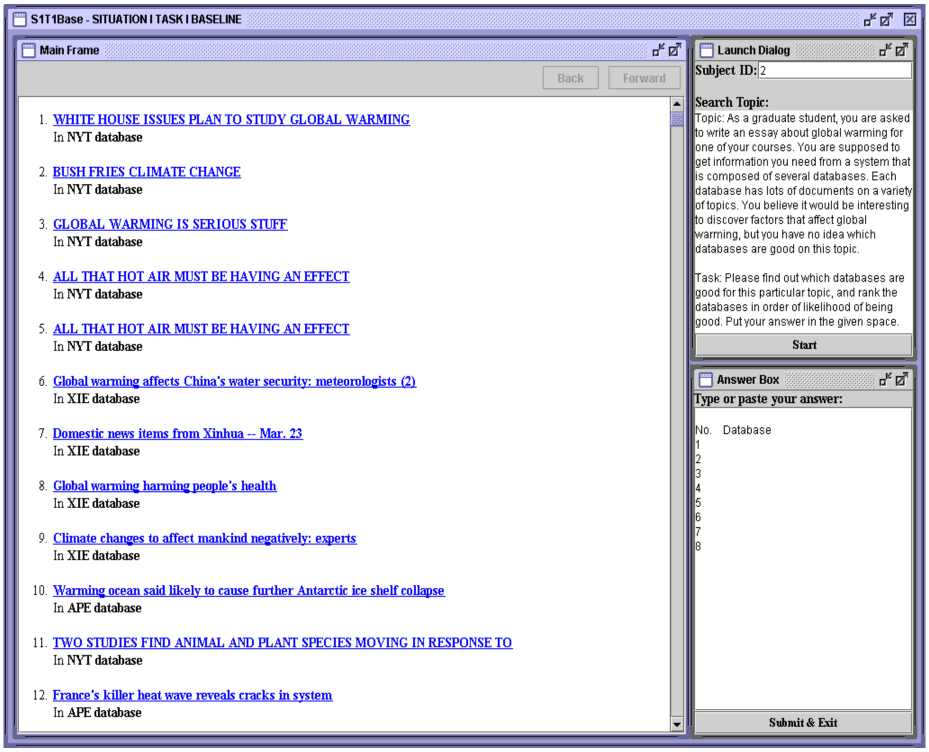


Figure 4.2 Baseline system B1.1 (Situation 1, Task 1)

4.4.2 Situation 1 (Scanning), Task 2 (T1.2, Find comments from an electronic book)

T1.2: A person is in the process of preparing an address for a conference. S/he recalls some germane comments from a known electronic book but cannot remember the exact wording of the comments. S/he needs to find out the exact quotations.

Description: Since the person has only a general idea about the quotations, s/he needs to *scan* through the *meta-information* to generate some candidate quotation page numbers. This person might look initially at the table of contents of the book for places where the quotations might occur. Then s/he goes to those pages and scans through them roughly to see if the desired quotations are there and, if so, record the quotations.

In this task, a combination of such IR support techniques as table of contents navigation visualization, and scrolling are used to support the scanning strategy. The interactions between the user and the system are related to the situation, task and goal.

4.4.2.1 System Design (E1.2/B1.2)

The goal is to test whether table of contents navigation would guide the user and help get the desired result more effectively. In the experimental system (see Figure 4.3), the screen is split into two parts. The left side is the table of contents of the electronic book. When the user clicks on a chapter, that chapter will be shown in greater detail with each section's title being displayed. For the chapter that the user is currently looking at on the left, clicking on the section title leads to a display of the section content in the right side of the screen. In the baseline system (see Figure 4.4), each paragraph of the book is indexed as a separate document. A ranked list of paragraphs retrieved with respect to the topic is provided, with each paragraph represented by a brief summarization consisting of the first sentence of that paragraph. After clicking on the summarization, the whole paragraph will be shown in a new screen.

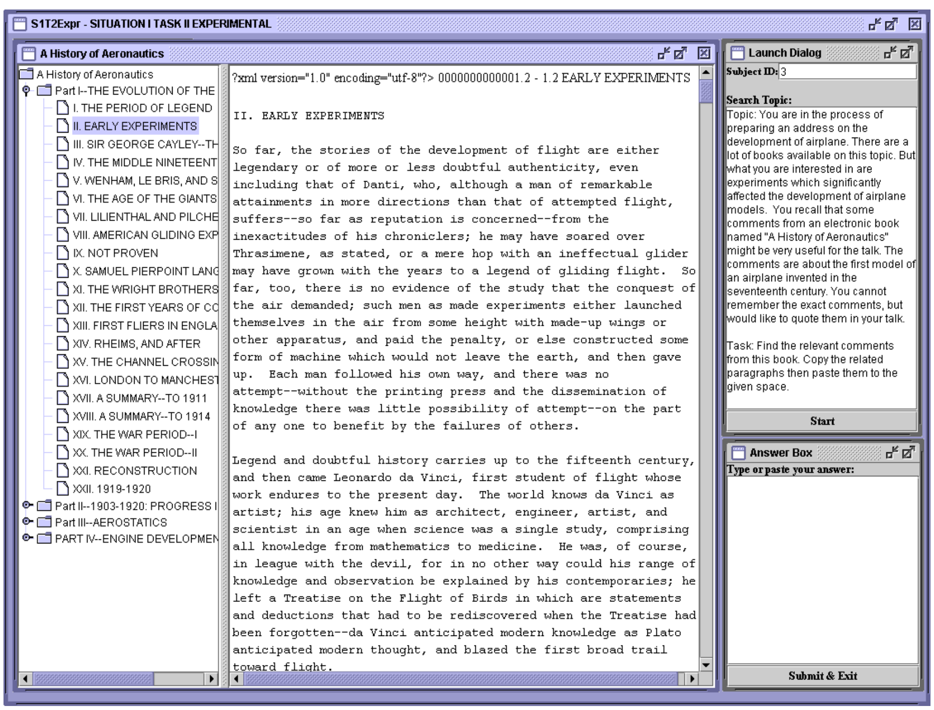


Figure 4.3 Experimental system E1.2 (Situation 1, Task 2)

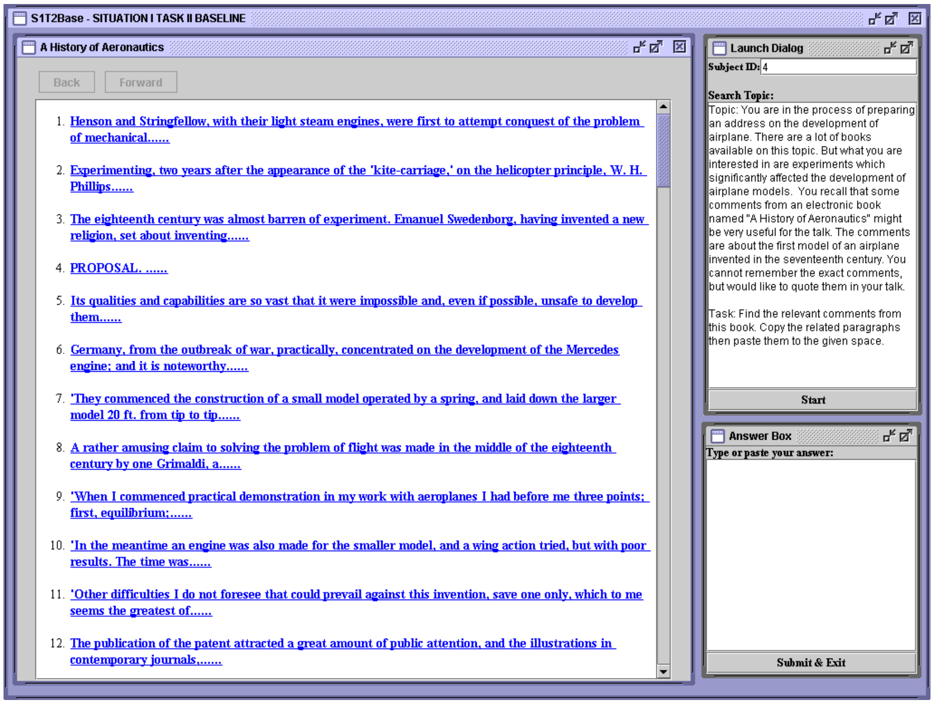


Figure 4.4 Baseline system B1.2 (Situation 1, Task 2)

4.4.3 Situation 2 (Searching), Task 1 (T2.1, Find relevant documents)

T2.1: A person is interested in one particular topic. S/he wants to find some good documents on this topic from a database.

Description: The person needs to construct a *systematic search* within one database for the particular topic in order to identify documents of interest.

The person needs to formulate a query based on the given task. The query would be compared using a best match technique against the index terms associated with chosen databases. The results of the query can be represented by clustering because clustering shows the relationship between documents, as well as the relationship between terms in the clusters and query terms or other terms that might turn out to be useful. Query-based clusters (that is, clusters reflecting the information problem expressed in a query) would be preferred because it is known that clustered displays based on topicality are useful for helping people find relevant information (Jardine & van Rijsbergen, 1971; Muresan, 2002). It is also believed that clusters can tell a person what the relationship is between different clusters in the specific database at a glance. To accomplish this, the system needs to get the retrieval results and cluster them. (Note: it doesn't need to cluster the entire database.) Each cluster has a short summary giving the centrality of that topic in the cluster, the number of documents in each cluster, and the number of documents likely to be relevant to the particular topic. Now the person can decide which clusters are desirable and then drill down to relevant documents within the clusters.

In this task, a combination of such IR support techniques as best match, clustering, clustered retrieval results display and following links are used to support searching. The interactions between the user and the system are related to the situation, task and goal.

4.4.3.1 System Design (E2.1/B2.1)

The aim is to test the difference in efficacy between the clustered retrieval results and a traditional "flat" ranked list. In both experimental (see Figure 4.5) and baseline systems (see Figure 4.6), there are a query box and a search button. Only one database, NYT, is used in this task. In the experimental system, the experimenter types in a query for the user, and the search button is disabled after the search. The retrieved results are shown by clusters, and the related labels and the snippets of several documents are displayed for each cluster. These clusters are ranked based on the posting frequency of the query terms. The baseline system provides a ranked list of the retrieved documents.

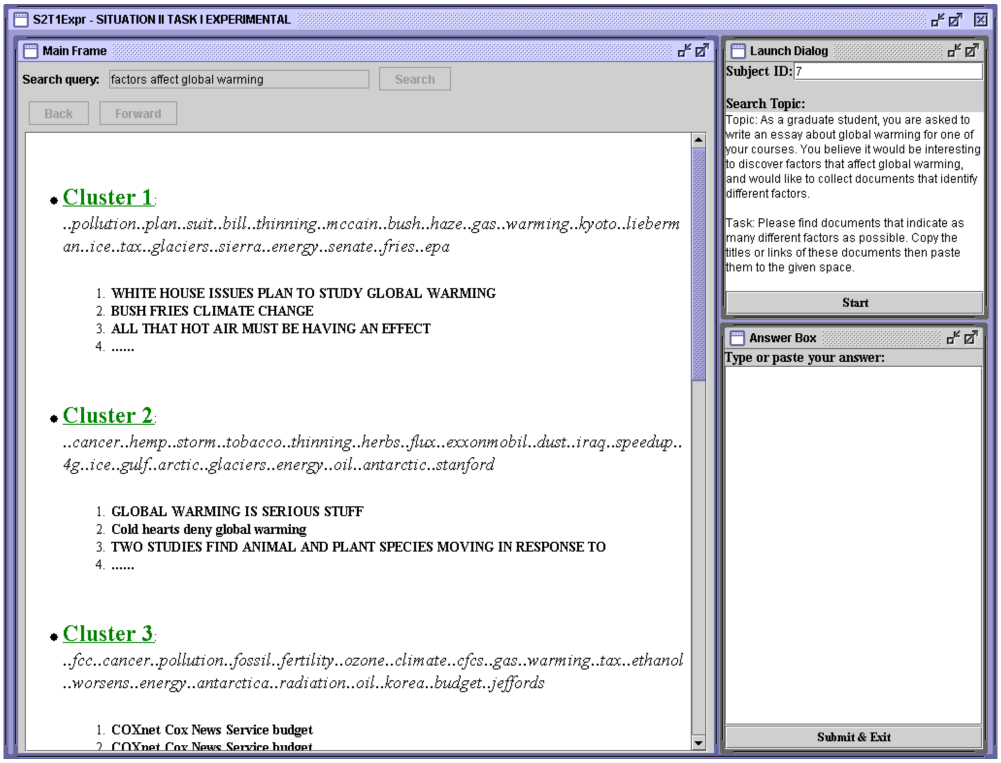


Figure 4.5 Experimental system E2.1 (Situation 2, Task 1)

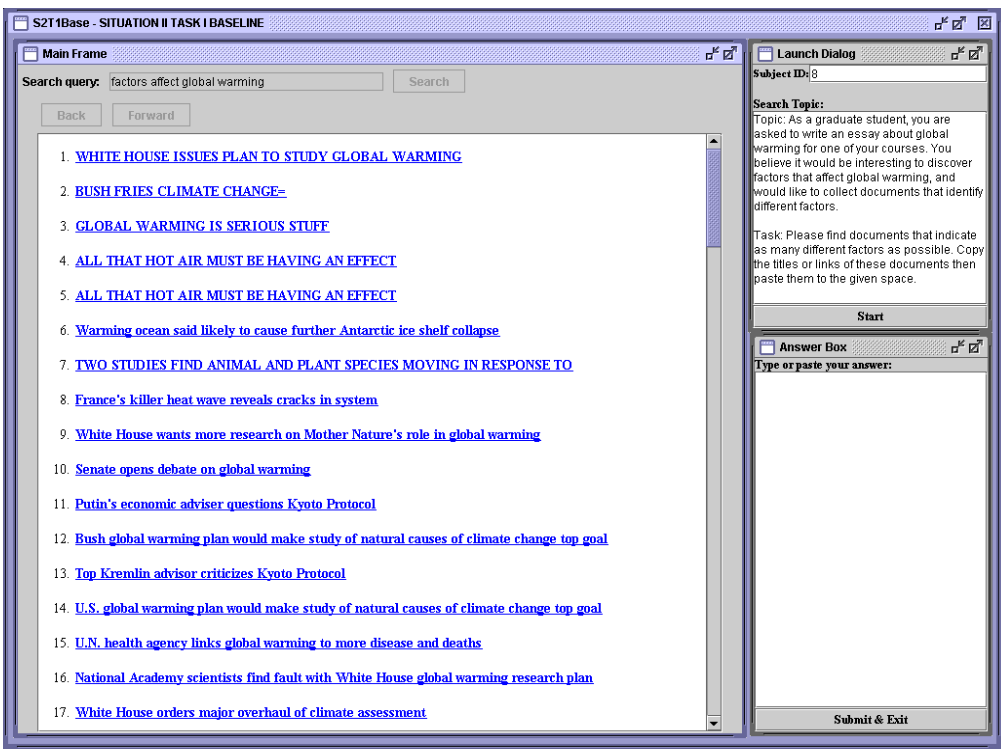


Figure 4.6 Baseline system B2.1 (Situation 2, Task 1)

4.4.4 Situation 2 (Searching), Task 2 (T2.2, Find the name of an electronic book)

T2.2: A person is preparing an address for a conference. S/he recalls that a certain electronic book might be helpful. But s/he cannot exactly remember the name of the book.

Description: This person has a vague recollection about a book that s/he saw. S/he needs to improve her/his knowledge of some characteristics of the book, such as author, title and publication year. Thus, s/he might need to *search* the system on terminological fragments of those data elements. In this situation, it might be helpful to give the person an opportunity to see information according to such characteristics or data elements. The items in the database, catalog or electronic book could be indexed to support a best match or exact match technique within different fields such as title, author, publication year, publisher,and publication place. The retrieved results will be displayed as a list of complete citations of the books. Then the person can see the table of contents of each book by following the link of each citation.

In this task, a combination of such IR support techniques as best/exact match, indexing, fielded query search and following links are used to support searching. The interactions between the user and the system are related to the situation, task and goal.

4.4.4.1 System Design (E2.2/B2.2)

The goal is to test the difference between fielded query search and generic search. In the experimental system (see Figure 4.7), several fields such as title, publication year, publisher, author and publication place are provided to help the user search the available books. The retrieved results are displayed as complete citations of the books. The table of contents is shown at the bottom after clicking each citation. In the baseline system (see Figure 4.8), only a single all-fields search box is provided and the retrieved results are also displayed as complete citations of the books, and the table of contents of each book is displayed after clicking each citation.

For the experimental system, the books are retrieved using LEMUR's Indri structured query. The query is an "AND" combination of all the input fields. For the fields "author", "publication year", "publisher" and "place", it needs to be an exact match. The retrieved results are grouped by publication years.

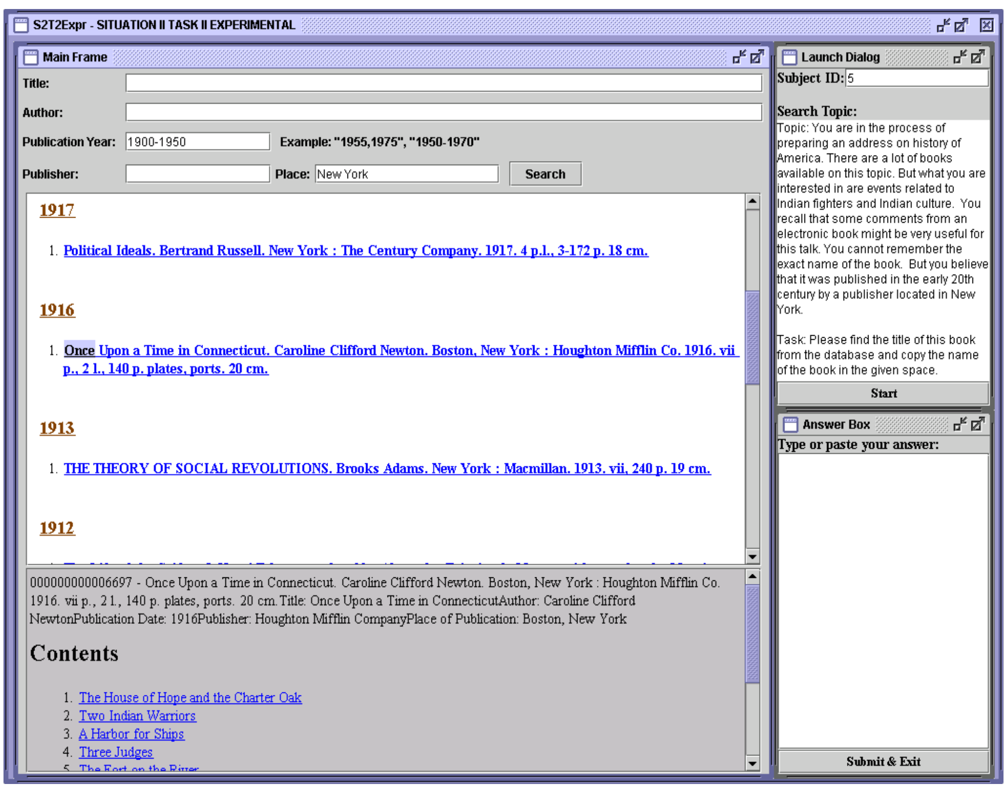


Figure 4.7 Experimental system E2.2 (Situation 2, Task 2)

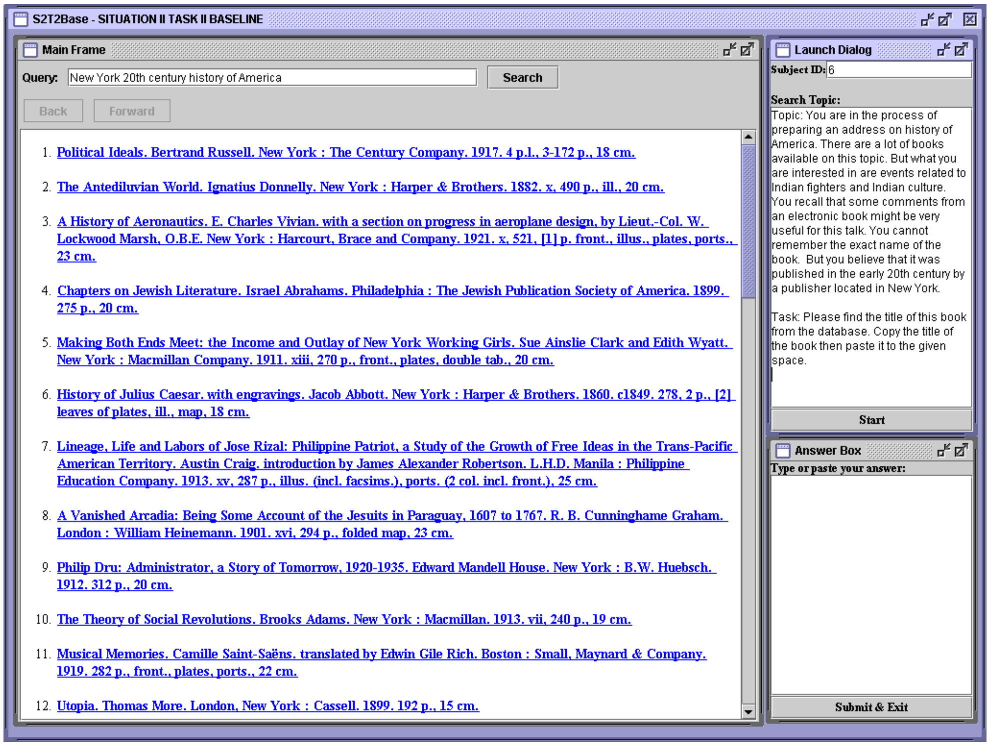


Figure 4.8 Baseline system B2.2 (Situation 2, Task 2)

4.5 Tasks and Topics

Four topics per task are provided for each set of experiments. These topics are designed to motivate scanning or searching accordingly.

These experiments try to ensure that the tasks are close to real world situations. Borlund and Ingwersen (1997) pointed out that the experimental settings of current laboratory experiments are unrealistic. They further proposed a simulated work task situation in which search scenarios reflecting real-life situations were described. Borlund (2000) found that these scenarios created the same behavior as real information needs. In this study, the simulated work task situation model was used to make subjects' behavior as real as possible, and thereby hope to get more robust results.

Following are the tasks and topics used in this experiment.

Situation 1 - Task 1 (Scanning, Identify best databases)

1. Topic: As a graduate student, you are asked to write an essay about global warming for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of documents on a variety of topics. You believe it would be interesting to discover factors that affect global warming, but you have no idea which databases are good on this topic.

Task: Please find out which databases are good for this particular topic, and rank the databases in order of likelihood of being good. Put your answer in the given space.

2. Topic: As a graduate student, you are asked to write an essay about air pollution for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of documents on a variety of topics. You believe it would be interesting to discover factors that cause air pollution, but you have no idea which databases are good on this topic.

Task: Please find out which databases are good for this particular topic, and rank the databases in order of likelihood of being good. Put your answer in the given space.

3. Topic: As a graduate student, you are asked to write an essay about high blood pressure for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of documents on a variety of topics. You believe it would be interesting to discover methods that reduce high blood pressure, but you have no idea which databases are good on this topic.

Task: Please find out which databases are good for this particular topic, and rank the databases in order of likelihood of being good. Put your answer in the given space.

4. Topic: As a graduate student, you are asked to write an essay about international trade for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of documents on a variety of topics. You believe it would be interesting to discover factors that affect international trade in cotton, but you have no idea which databases are good on this topic.

Task: Please find out which databases are good for this particular topic, and rank the databases in order of likelihood of being good. Put your answer in the given space.

Situation 1 - Task 2 (Scanning, Find comments from an electronic book)

1. Topic: You are in the process of preparing an address on the development of airplane. There are a lot of books available on this topic. But what you are interested in are experiments which significantly affected the development of airplane models. You recall that some comments from an electronic book named "A History of Aeronautics" might be very useful for the talk. The comments are about the first model of an airplane invented in the seventeenth century. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Find the relevant comments from this book. Copy the related paragraphs then paste them to the given space.

2. Topic: You are in the process of preparing an address on history of America. There are a lot of books available on this topic. But what you are interested in are events related to Indian fighters and Indian culture. You recall that some comments from an electronic book named "Once upon a time in Connecticut" might be very useful for this talk. It is about two Indian warriors, Uncas and Miantonomo, and the comments tell the

story of the fate of Miantonomo. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Find the relevant comments from this book. Copy the related paragraphs then paste them to the given space.

3. Topic: You are in the process of preparing an address on childhood education. There are a lot of books available on this topic. But what you are interested in is the history of censorship of books for kids. You recall that some comments from an electronic book named “Report of the Special Committee on Moral Delinquency in Children and Adolescents The Mazengarb Report (1954)” might be very useful for this talk. The comments talked about what kinds of publications children should not read. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Find the relevant comments from this book. Copy the related paragraphs then paste them to the given space.

4. Topic: You are in the process of preparing an address on business. There are a lot of books available on this topic. But what you are interested in is the development of the domestic bird business. You recall that some data from an electronic book named “The Dollar Hen” might be very useful for this talk. The data are about the development of the poultry industry in different states in USA. You cannot remember the exact data, but would like to quote them in your talk.

Task: Find the relevant data from this book. Copy the data then paste them to the given space.

Situation 2 - Task 1 (Searching, Find relevant documents)

1. Topic: As a graduate student, you are asked to write an essay about global warming for one of your courses. You believe it would be interesting to discover factors that affect global warming, and would like to collect documents that identify different factors.

Task: Please find documents that indicate as many different factors as possible. Copy the titles or links of these documents then paste them to the given space.

2. Topic: As a graduate student, you are asked to write an essay about air pollution for one of your courses. You believe it would be interesting to discover factors that cause air pollution, and would like to collect documents that identify different factors.

Task: Please find documents that indicate as many different factors as possible. Copy the titles or links of these documents then paste them to the given space.

3. Topic: As a graduate student, you are asked to write an essay about high blood pressure for one of your courses. You believe it would be interesting to discover methods that reduce high blood pressure, and would like to collect documents that identify different methods.

Task: Please find documents that indicate as many different methods as possible. Copy the titles or links of these documents then paste them to the given space.

4. Topic: As a graduate student, you are asked to write an essay about international trade for one of your courses. You believe it would be interesting to discover factors that affect international trade in cotton, and would like to collect documents that identify different factors.

Task: Please find documents that indicate as many different factors as possible. Copy the titles or links of these documents then paste them to the given space.

Situation 2- Task 2 (Searching, Find the name of an electronic book)

1. Topic: You are in the process of preparing an address on the development of airplanes. There are a lot of books available on this topic. But what you are interested in are experiments which significantly affected the development of airplane models. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you remember that it is written by Chares Vian, or someone like that, and was published in the early 20th century.

Task: Please find the title of this book from the database. Copy the title of the book then paste it to the given space.

2. Topic: You are in the process of preparing an address on history of America. There are a lot of books available on this topic. But what you are interested in are events related to Indian fighters and Indian culture. You recall that some comments from an electronic book might be very useful for this talk. You cannot remember the exact name of the book. But you believe that it was published in the early 20th century by a publisher located in New York.

Task: Please find the title of this book from the database. Copy the title of the book then paste it to the given space.

3. Topic: You are in the process of preparing an address on childhood education. There are a lot of books available on this topic. But what you are interested in is the history of censorship of books for kids. You recall that some comments from an

electronic book might be very useful for this talk. You cannot remember the exact name of the book. But you believe that it was published in the 20th century.

Task: Please find the title of this book from the database. Copy the title of the book then paste it to the given space.

4. Topic: You are in the process of preparing an address on business. There are a lot of books available on this topic. But what you are interested in is the development of the domestic bird business. You recall that some data from an electronic book might be very useful for this talk. You cannot remember the exact name of the book. But you remember that it is written by Hatings, or someone like that, and was published in the early 20th century.

Task: Please find the title of this book from the database. Copy the title of the book then paste it to the given space.

4.6 Text Collections

There are two text collections (see Table 4.3): Collection 1 (HARD 2004 (Allan, 2005)) has several databases suitable for situation 1 - task 1 and situation 2 - task 1, while Collection 2 (50 books downloaded from Project Gutenberg (http://www.gutenberg.org/)) has a book database which is suitable for situation 1- task 2 and situation 2 - task 2.

Table 4.3

Two Text Collections

Collections	Situations	Tasks
TREC HARD 2004	Scanning	1.Identify best databases
	Searching	1.Find relevant documents
Fifty books from Project Gutenberg	Scanning	2.Find comments from an electronic book
	Searching	2.Find the name of an electronic book

4.6.1 Collection 1

The HARD (High Accuracy Retrieval from Documents) project in TREC (Text Retrieval Conference) aims to discover methods to improve the search result accuracy of IR systems by taking into account additional information about the searcher and the search context (Allan, 2005). The current study needs to have a collection of several different databases, and the HARD 2004 collection fits this need well for the following three reasons. First, as a group member of the Rutgers HARD Track project, the author has experience using the collection for several years. Second, it is free and convenient. Third, documents in this collection are collected and distributed by professionals.

The HARD 2004 evaluation uses the HARD 2004 English newswire corpus, which was collected and distributed by the Linguistic Data Consortium for the HARD project. This corpus includes one year (2003) of newswire data, from eight sources: AFP (Agence France Press), APW (Associated Press), CNA (Central News Agency), LAT (Los Angeles Times/Washington Post), NYT (New York Times), SLN (Salon.com), UMM (Ummah Press), and XIN (Xinhua News Agency – English). The documents have been cleaned and standardized by the Linguistic Data Consortium. Each document was assigned a unique document ID with three-letter newswire source abbreviation, the year, month, day and chronological sequence of publication. A sample topic is given in Appendix A. Table 4.4 shows the sources and the number of documents for each source.

Table 4.4

Structure of HARD 2004 Corpus (after Allan, 2005)

Newswires	No. of documents	Size (Mbs)
AFP	226,777	497
APW	236,735	644
CNA	4,011	6
LAT	34,145	107
NYT	27,835	105
SLN	3,134	28
UMM	2,557	5
XIN	117,516	183
TOTAL	**652,710**	**1,575**

4.6.2 Collection 2

There exist several online book projects, such as the Million Book Project (http://www.archive.org/details/millionbooks), Open Source Books (http://www.archive.org/details/opensource), and Project Gutenberg. Since this study needs a database in which books are all structured so that LEMUR can index them, Project Gutenberg was chosen.

Project Gutenberg is the oldest producer of free electronic books (eBooks or eTexts) on the Internet. It has a collection of more than 15,000 eBooks. Most of the eBooks are older literary works that are in the public domain in USA. These books can be freely downloaded, read, and redistributed for non-commercial use, and many books are in HTML format which can be easily changed to suit our study.

The books downloaded for the current study are non-fictional books in English and HTML format and have a table of contents. There are 50 books in collection 2. The complete citation of each book was retrieved from OCLC Connexion Integrated Cataloging and Metadata Services (OCLC http:// http://connexion.oclc.org/). Each citation includes title, author, publisher, publication place, publication year, and pagination.

4.7 Experimental Design

This experiment is a within-subjects design. Subjects conducted several searches on different topics that are suitable for scanning or searching. Each subject searched half of the topics using E1.1 (database summary), E1.2 (table of contents navigation), E2.1 (clustered retrieval results), or E2.2 (fielded query), and half using B1.1, or B1.2, or B2.1, or B2.2. Then the experiment was repeated with exchanging the order of the systems. Within the topic block the topic order was randomly assigned. No two subjects used the same order of topics and the same order of systems.

4.8 Sampling

Subjects were mainly recruited from Rutgers graduate students. The recruitment notice was posted in various Rutgers departmental listservs, and was announced in class. The search sessions were held at the usability lab of Rutgers SCILS building.

4.9 Measures and Variables

This study chose *user satisfaction*, *result correctness* and *aspectual recall* to measure search performance, as well as measures of effort such as *time to complete a task*, and *degree of interaction* in conducting the search.

User satisfaction is one of the most popular performance measures (Harter & Hert, 1997). Result correctness was judged by the assessor. In a question-answering task environment, user satisfaction and result correctness have been widely accepted as important factors indicating users' perception of retrieval effectiveness (Belkin et al., 2001; Belkin et al., 2002; Belkin et al., 2003). User satisfaction is measured by asking each subject to rate his or her own satisfaction with the search results on a 7-point scale ranging from *Not at all* to *Extremely*. Result correctness is measured as the assessor's rating of the saved book/paragraphs which answer the search topic on a 3-point scale: *Incorrect, Partially Correct,* and *Correct*. If the saved book is exactly the right book, or the saved paragraphs are exactly the right paragraphs that answer the search question, it is rated as "Correct." If the saved book is not exactly the right book, it is rated as "Incorrect." If the saved paragraphs only contain closely related paragraphs, it is rated as "Partially correct." Otherwise, it is rated as "Incorrect."

In the experiment for E2.1 (clustered retrieval results)/B2.1, aspectual recall was adopted as one of the measures of system performance because of the nature of the tasks (asking the subject to identify as many factors/methods of a topic as possible). Aspectual recall, a measure developed in the TREC Interactive Track (Dumais & Belkin, 2005) is the ratio of identified aspects to total aspects of the topics that are covered by the pooled submitted documents. Different subjects may use different wordings for similar aspects.

The assessor interpreted the aspects identified by the subjects, and grouped them into broader categories. In this study, the assessor is the experimenter.

Some interaction measures such as the number of iterations (*number of queries issued in a search*), number of final saved documents, number of documents/books viewed were also of interest. These measures have been shown to be valid in many experiments (Belkin et al., 2001; Belkin et al., 2002; Belkin et al., 2003; White, Ruthven, & Jose, 2003) and they help us to understand subjects' information-seeking activities.

4.10 Data Collection

In the experiment, an *entry questionnaire* (Appendix B(2)) gathered demographic and other background information. A *pre-search questionnaire* (Appendix B(3)) collected information about subjects' previous knowledge of the topic. A *post-search questionnaire* (Appendix B(4)) elicited opinions about the particular search. A *post-system questionnaire* (Appendix B(5)) collected opinions about the specific system. An *exit questionnaire* (Appendix B(6)) elicited opinions about the systems and the whole experimental process. The computer logged subjects' search activities (e.g., iterations, query input, time of task completion)**.**

4.11 Procedure

When subjects arrived, they completed an informed consent form (Appendix B(1)), which included detailed instructions about the experiment, and then the entry questionnaire. Next, they began the search on the first topic. For each topic, they filled out a pre-search questionnaire, then conducted the search and saved the answers in the given space. When they felt that a satisfactory answer was saved, or they ran out of time (each search was limited to 10 minutes), they went on to the next topic. Upon completion of each topic, they answered a brief post-search questionnaire. After completing the first two topics on one system, they filled out a post-system questionnaire. The same procedure was followed for the next set of topics using the second system, after which the exit questionnaire was given. This procedure is displayed in Figure 4.9.

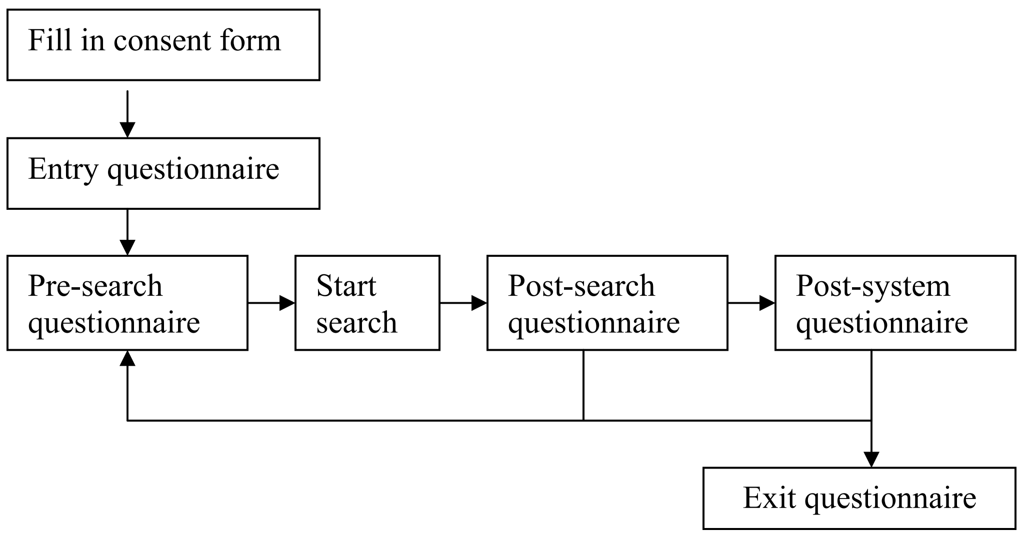


Figure 4.9 Experimental procedure (Experiment I)

Chapter 5

Research Problem 1: Results

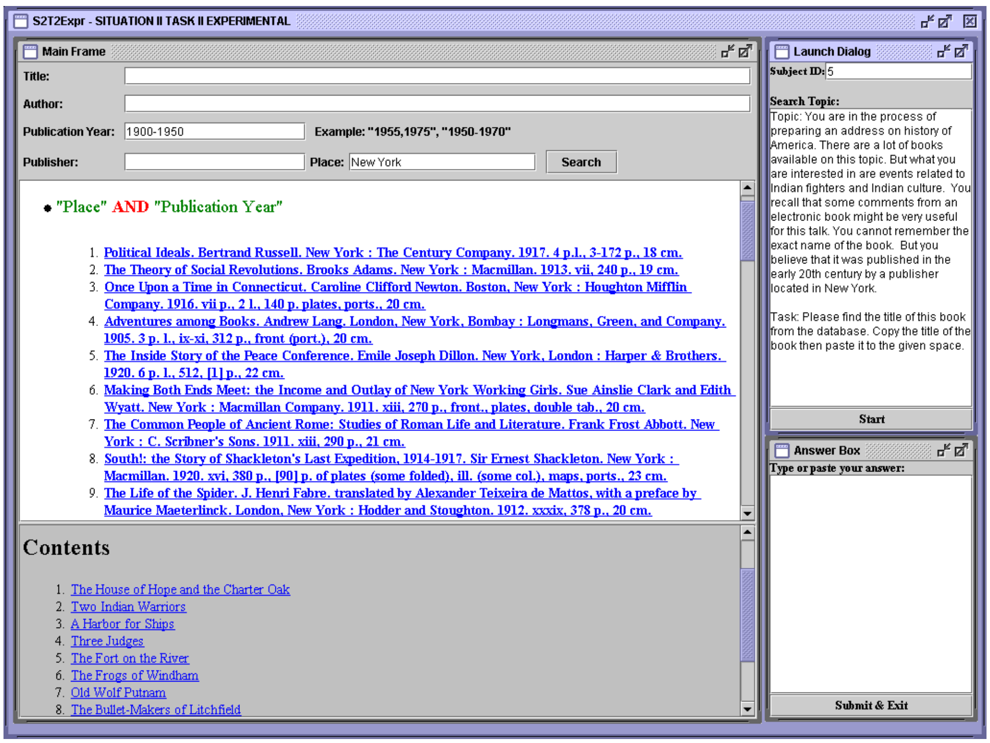
In this chapter, the results from the experiment I are presented. It starts with the findings and discussions of the pilot study. Then the subjects' characteristics and their computer and searching experience are presented. A description of the performance and interaction measures of the systems follows. Next, the data from the pre-search, post-search and post-system questionnaires is presented. Finally, the results from the exit questionnaires are reported.

5.1 Pilot Results of Experiment I

In the pilot study, two subjects were recruited to run each set of systems. Thus a total of eight subjects, who were all Rutgers graduate students, participated in this study.

5.1.1 Systems and Questionnaires

The subjects had no problems understanding the topics and tasks they needed to complete. Also, the systems and questionnaires proved to be valid based on the users' responses to the questions and the pilot results. Only a few changes had to be made. Firstly, some changes needed to be made to the questionnaires. For instance, two users suggested another question at the end of exit questionnaire, that is, "Do you have any other comments or suggestions?", so that they could provide more input about the systems. The other change was to clarify the fifth question of post-search questionnaire, "Did your previous knowledge help you?" One user was confused with the meaning of "knowledge": knowledge of searching or knowledge of the topic. This question was clarified as: "Did your previous knowledge of the topic help you?" in the future experiments. Secondly, in system E1.2 (table of contents navigation), some metadata were shown at the top of each section of the books after the user clicked on the related links. These metadata were later deleted to avoid confusion. Thirdly, it was noted that different Boolean combinations of fielded queries should be considered in system E2.2 (fielded query). For example, if the query used publication year and publication place, then the possible combinations could be: year AND place, year but NOT place, and place but NOT year. This can be seen in Figure 5.1 and Figure 5.2. Lastly, there were two dead document links that had to be fixed.



<u>Figure 5.1</u> Modified system E2.2 (Fielded query) part 1

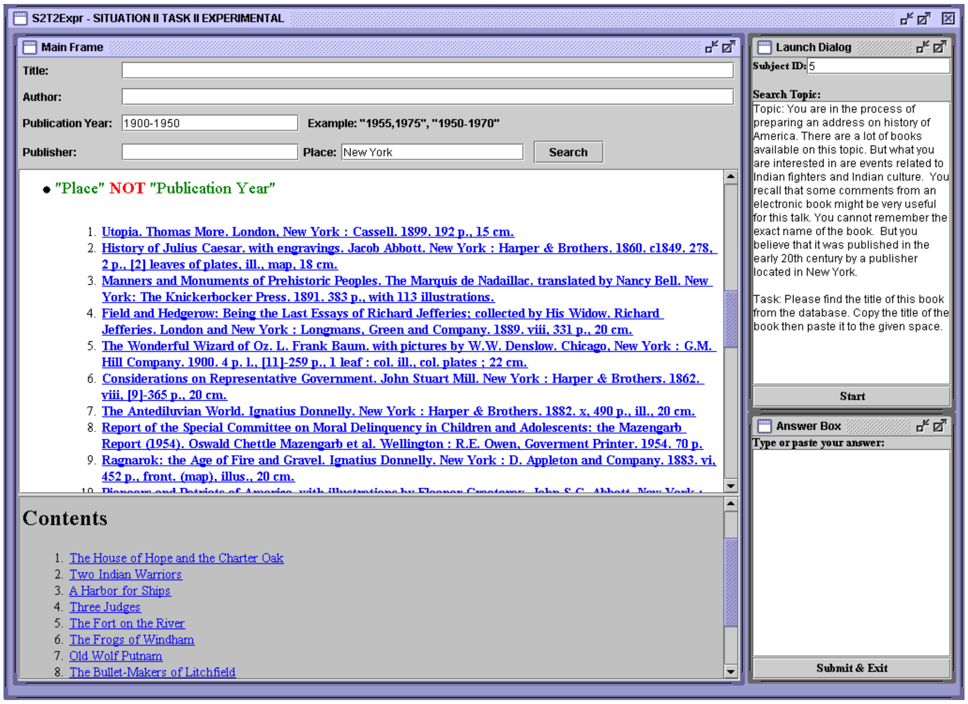


Figure 5.2 Modified system E2.2 (Fielded query) part 2

5.1.2 Preliminary Findings

Time, user satisfaction, and result correctness were used to measure the performance of all the systems (see Section 4.9 of Chapter 4 for more details). Results indicated that subjects spent less time and felt more satisfied using the experimental systems than using the baseline systems. For E1.2/B1.2 (table of contents navigation) and E2.2/B2.2 (fielded query), subjects found more correct answers using the experimental systems than using the baseline systems.

After further analyzing the questionnaire data, it was found that subjects showed strong preferences for the experimental systems. More specifically, subjects liked such features as database summary in system E1.1, table of contents navigation in system E1.2, clustered retrieval results display in system E2.1, and fielded query search in system E2.2. Subjects indicated that the experimental systems were easier to learn, easier to use, more understandable, and more useful than the baseline systems. Results also showed that the experimental systems made subjects feel that it was easier to start the search and easier to search the topic, and feel that they have more time to do the search than the baseline systems.

In summary, the pilot results supported the hypotheses. Although the results were not generalizable due to the small sample, they encouraged us to continue the study with the succeeding experiments.

5.2 Results of Experiment I

5.2.1 Subjects

Thirty-two Rutgers graduate students (excluding the eight pilot subjects) participated the experiment. Sixteen (50%) were female and sixteen (50%) were male. Seventy-five percent of the subjects were between 26-35 years of age, while others (25%) range from 36-65 years of age. About half (46.9%) of them were in the library and information studies field, among which about one third had a master's degree and were in the Ph.D. program (see Table 5.1).

Table 5.1

Subject Characteristics (Experiment I)

Characteristics	Values	No. of subjects
Age	26-35	24
	36-45	3
	46-55	4
	56-65	1
Current major	Library and information studies	15
	Computer science	5
	Mechanical engineering	2
	Mathematics	2
	Communication	2
	Others	6
Degree earned	Ph.D.	
	Master	18
	Bachelor	14

Subjects' searching experience were measured on the entry questionnaire using a 7-point scale, where 1 = "none"; 4 = "some"; and 7 = "a great deal." Subjects were asked to indicate their level of expertise with computers and with online searching on a 7-point scale, where 1 = "novice" and 7 = "expert." Subjects were also asked about the frequency of their computer use and searching on a 7-point scale, where 1 = "never"; 4 = "monthly"; and 7 = "daily." Subjects were asked to indicate whether they could usually find what they were looking for on a 7-point scale, where 1 = "rarely"; 4 = "sometimes"; and 7 = "often." These results are listed below in Table 5.2. Subjects had very frequent use of computers (M=7.00, SD=0), high expertise of computers (M=5.91, SD=1.06), high searching experience of catalogs (M=5.69, SD=1.4) and WWW (M=6.81, SD=0.47), very high frequency of search (M=6.50, SD=0.84), high expertise of searching (M=5.88, SD=0.83), and very high confidence in finding what they need from searching (M=6.34, SD=0.79). Subjects also had long-term experience in searching (M=8.75 years, SD=4.92 years). However, their searching experience of commercial systems was relatively low (M=3.84, SD=1.85).

Table 5.2

Computer and Searching Experience of Subjects (Experiment I)

Demographic data	Mean (standard deviation)
Computer daily use	7.00 (0)
Expertise of computer	5.91 (1.06)
Searching experience of Catalog	5.69 (1.4)
Searching experience of commercial systems	3.84 (1.85)
Searching experience of WWW	6.81 (0.47)
Frequency of search	6.50 (0.84)
Search information found	6.34 (0.79)
Expertise of searching_	5.88 (0.83)
Number of years of searching experience	8.75 (4.92)

5.2.2 Performance

Time of task completion, user satisfaction, result correctness, and aspectual recall were the measures of performance in this experiment. Time was collected by system logs. User satisfaction was assessed by post-search questionnaires. Subjects were asked to rate their satisfaction with the search results on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely." Aspectual recall was calculated based on aspects identified by the assessor. In this study, the experimenter was the assessor.

Specifically, time and user satisfaction were performance measures for all systems. Result correctness was the performance measure of system E1.2 (table of contents navigation) / B1.2 and E2.2 (fielded query) / B2.2. Aspectual recall was the performance measure of system E2.1 (clustered retrieval results) / B2.1. Table 5.3

summarizes the mean and standard deviation values of these measures for each system. SPSS 14.0 was used to run the data analysis.

Table 5.3

Performance of Systems (Experiment I)

Systems	Mean (Standard deviation)			
	Time (mins)	Result satisfaction (1-7)	Result correctness (0-2)	Aspectual recall
B1.1	8.32 (1.97)	4.19(1.11)		
E1.1	7.26 (1.37)	4.81 (1.05)		
B1.2	7.66 (2.32)	4.56 (1.55)	0.88 (0.96)	
E1.2	5.56 (1.88)	5.63 (0.81)	1.19 (0.91)	
B2.1	9.20 (1.25)	4.00 (2.00)		0.56 (0.18)
E2.1	8.71 (1.60)	5.19 (1.52)		0.63 (0.16)
B2.2	5.39 (2.36)	4.06 (1.95)	1.25 (1.00)	
E2.2	3.20 (1.41)	5.50 (1.67)	1.63 (0.81)	

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

Tables 5.3 and 5.4 show results from ANOVA that indicate that subjects using E1.1 (database summary) spent less time (M =7.26, SD=1.37) than those using B1.1 (M =8.32, SD=1.97), although the difference was not significant, $F(1,30)=3.09$, $p=0.090$. Subjects using E1.2 (table of contents navigation) spent significantly less time (M =5.56, SD=1.88) than those using B1.2 (M =7.66, SD=2.32), $F(1,30) =7.866$, $p=0.009$. Subjects using E2.1 (clustered retrieval results) spent less time (M=8.71, SD=1.60) than those using B2.1 (M=9.20, SD=1.25), although not significantly so, $F(1,30)=0.9$, $p=0.350$. Subjects using E2.2 (fielded query) spent significantly less time (M =3.20, SD=1.41) than those using B2.2 (M =5.39, SD=2.63), $F(1,30)=10.183$, $p=0.003$.

Pearson chi-square test showed that there was no significant relationship between system and result correctness, although subjects using E1.2 (table of contents navigation) found more correct answers (M =1.19, SD=0.91) than those using B1.2 (M =0.88, SD=0.96), χ^2 =1.178, df=2, p=0.555. Subjects using E2.2 (fielded query) found more correct answers (M =1.63, SD=0.81) than subjects using B2.2 (M =1.25, SD=1.00), χ^2=1.340, df=2, p=0.518. Table 5.5 describes the distribution of answer correctness across the systems. Subjects using the experimental systems (E1.2 (table of contents navigation) or E2.2 (fielded query)) got more correct answers and fewer incorrect answers than those using the baseline systems (B1.2 or B2.2), but not significantly so.

Wilcoxon signed rank test results showed that subjects felt more satisfied with the results using the experimental systems than the baseline systems. More specifically, subjects using E1.1 (database summary) felt more satisfied (M = 4.81, SD =1.05) than those using B1.1 (M =4.19, SD =1.11), although no significant results were found, Z=-1.398, p=0.162. Subjects using E1.2 (table of contents navigation) felt significantly more satisfied (M =5.63, SD =0.81) than those using B1.2 (M =4.56, SD =1.55), Z=-2.738, p=0.006. Subjects using E2.1 (clustered retrieval results) felt more satisfied (M =5.19, SD =1.52) than those using B2.1 (M =4.00, SD =2.00), although not significantly so, Z=-1.283, p=0.199. Subjects using E2.2 (fielded query) felt more satisfied (M =5.50, SD =1.67) than those using B2.2 (M =4.06, SD =1.95), although no significant results were found, Z=-1.583, p=0.113.

Regarding aspectual recall for system E2.1 (clustered retrieval results)/B2.1, results showed that subjects using E2.1 (clustered retrieval results) found more relevant aspects (M =0.63, SD=0.16) than those using B2.1 (M =0.56, SD=0.18), although not significantly so, F(1,30)=1.319 , p=0.260.

Table 5.4

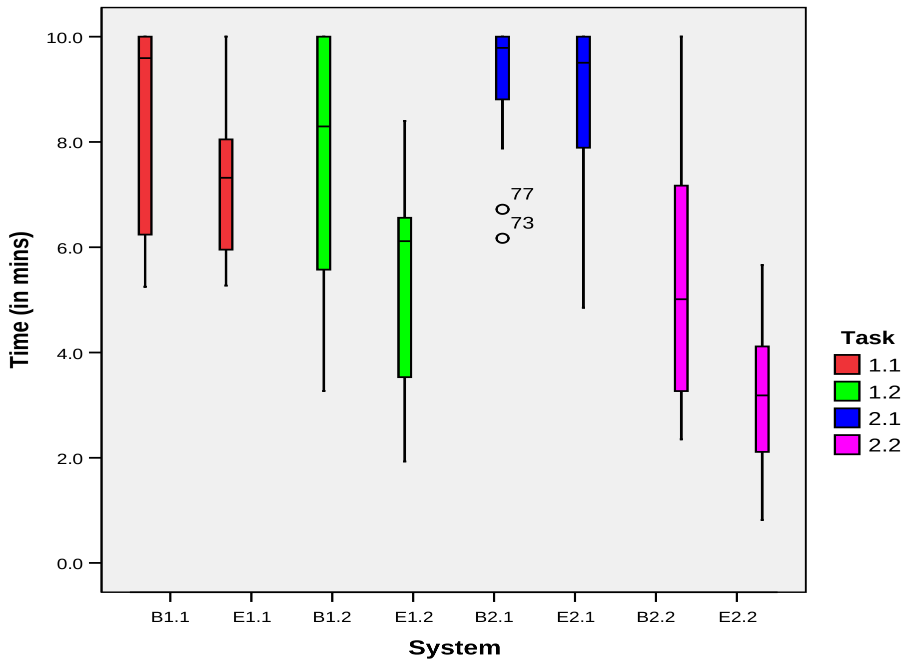
Significance Value of Systems (Experiment I)

Systems	ANOVA		Chi-square	Wilcoxon signed rank
	Time	Aspectual recall	Result correctness	Result satisfaction
B1.1	0.090			0.162
E1.1				
B1.2	0.009*		0.555	0.006*
E1.2				
B2.1	0.350	0.260		0.199
E2.1				
B2.2	0.003*		0.518	0.113
E2.2				

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

*$p < 0.01$

It should be noticed that subjects using the experimental systems (E1.1 (database summary) or E2.1 (clustered retrieval results) spent less time than those using the baseline systems (B1.1 or B2.1) though no significant results were found. Since time was also a very important measure for efficiency, a boxplot (see Figure 5.3) was used to show the distributions of time across all the systems. Subjects using system E1.1 (database summary)/B1.1 and E2.1 (clustered retrieval results)/B2.1 spent much more time than those using E1.2 (table of contents navigation)/B1.2 and E2.2 (fielded query)/B2.2.



Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

Figure 5.3 Time distributions across systems (Experiment I)

Table 5.5

Result Correctness across Systems (Experiment I)

Systems	Result correctness		
	Incorrect	Partially correct	Correct
B1.2	8	2	6
E1.2	5	3	8
B2.2	6		10
E2.2	3		13

Note. E1.2: table of contents navigation; E2.2: fielded query.

5.2.3 Interaction

Table 5.6 defines the interaction variables for the respective systems, including number of iterations, number of final saved documents, number of documents/books viewed.

Table 5.6

Variables Used to Describe Search Behavior of Interaction (Experiment I)

Variables	Definitions
Number of iterations	The total number of queries issued by the searcher during the entire search process
Number of final saved documents	The total number of documents which were saved by the searcher at the end of the search
Number of documents/books viewed	The total number of documents/books whose contents were displayed to the searcher

Table 5.7

Mean and Standard Deviation of Interaction Variables (Experiment I)

Interaction Measures	Systems							
	B1.1	E1.1	B1.2	E1.2	B2.1	E2.1	B2.2	E2.2
Number of iterations							4.69 (3.22)	4.06 (4.75)
Number of final saved documents					7.38 (3.07)	6.63 (2.25)		
Number of documents/ books viewed	4.13 (3.67)	5.31 (4.30)			14.19 (7.22)	12.50 (7.20)	5.63 (4.59)	5.44 (6.63)

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

In Table 5.7, ANOVA results showed that subjects using E2.2 (fielded query) had fewer iterations (M=4.06, SD =4.75) than those using B2.2 (M =4.69, SD =3.22), although not significantly so, $F(1,30)=0.19$, $p=0.666$. ANOVA results also showed that subjects using E2.1 (clustered retrieval results) finally saved fewer documents (M =6.63, SD =2.25) than those of B2.1 (M =7.38, SD =3.07), but not significantly so, $F(1,30)=0.621$, $p=0.437$. ANOVA results showed that subjects using E1.1 (database summary) viewed more documents (M =5.31, SD =4.30) than that of B1.1 (M =4.13, SD =3.67), but no significant results were found, $F(1,30)=0.892$, $p=0.354$. Subjects using E2.1 (clustered retrieval results) viewed fewer documents (M =12.50, SD =7.20) than those using B2.1 (M =14.19, SD =7.22), although the difference is not significant, $F(1,30)=0.438$, $p=0.513$. Subjects using E2.2 (fielded query) viewed fewer books (M =5.44, SD =6.63) than those using B2.2 (M =5.63, SD =4.59), although not significantly so, $F(1,30)=0.089$, $p=0.768$.

5.2.4 Pre-search Questionnaire

In the pre-search questionnaire, subjects were asked about their familiarity and expertise with the given topic on a 7-point scale, where 1 = “not at all”; 4 = “somewhat”; and 7 = “extremely.” Subjects were asked to indicate their level of expertise with the given topic on a 7-point scale, where 1 = “novice”; and 7 = “extremely.” Table 5.8 shows the mean and standard deviation of these two variables across the topics. Generally, subjects were more familiar and had more expertise with topics in situation 1- task 1 and situation 2 - task 1 than those of situation 1 - task 2 and situation 2 - task 2.

Table 5.8

Topic Familiarity and Expertise (Experiment I)

Tasks	Topic No.	Topics	Mean (standard deviation)	
			Topic familiarity	Topic expertise
Situation 1 - Task 1: identify best databases	1	Global warming	4.00 (1.31)	2.75 (1.16)
	2	Air pollution	3.75 (1.67)	2.63 (1.06)
	3	High blood pressure	3.38 (2.00)	2.50 (1.41)
	4	International trade in cotton	1.75 (0.71)	1.38 (0.74)
Situation 1 - Task 2: find comments from an electronic book	1	Development of airplane models	1.63 (0.74)	1.38 (0.74)
	2	History of America	1.38 (0.74)	1.50 (0.76)
	3	Childhood education	2.00 (1.41)	2.00 (1.41)
	4	Development of the domestic bird business	1.13 (0.35)	1.50 (1.07)
Situation 2 - Task 1: find relevant documents	1	Global warming	3.38 (1.85)	2.63 (1.30)
	2	Air pollution	3.25 (1.67)	2.38 (1.51)
	3	High blood pressure	3.00 (1.85)	2.50 (1.77)
	4	International trade in cotton	1.50 (0.93)	1.25 (0.46)
Situation 2 - Task 2: find the name of an electronic book	1	Development of airplane models	1.88 (1.36)	1.75 (1.39)
	2	History of America	1.75 (1.04)	1.63 (0.74)
	3	Childhood education	1.38 (0.52)	1.25 (0.46)
	4	Development of the domestic bird business	1.25 (0.46)	1.25 (0.46)

5.2.5 Post-search Questionnaire

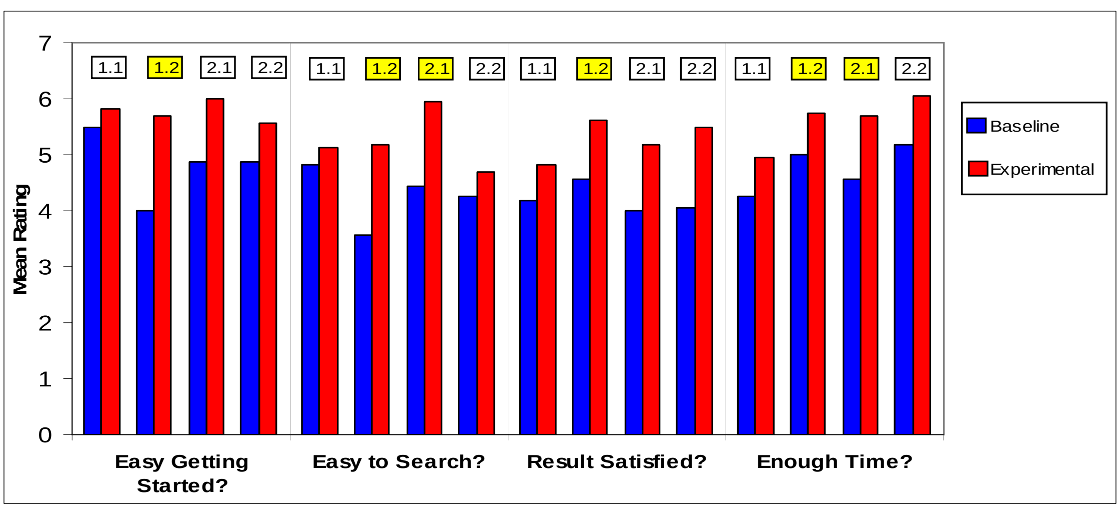
Subjects' opinions about each task were assessed by the post-search questionnaire. Subjects were asked whether it was easy to get started on the search, whether it was easy to do the search on the specific topic, whether they were satisfied with the results, and whether they had enough time on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely." Table 5.9 lists the mean values of each system for these questions. The systems were compared for each question based on each subject's responses using the non-parametric Wilcoxon signed rank test. SPSS 14.0 was used to do the analysis.

Results showed that it was easier to get started using E1.1 (database summary) (M=5.81, SD=0.91) than using B1.1 (M=5.50, SD=1.03), although not significantly so, Z=-1.155, p=0.248. It was significantly easier to get started using E1.2 (table of contents navigation) (M=5.69, SD=0.70) than using B1.2 (M=4.00, SD=1.86), Z=-3.028, p=0.002. It was easier to get started using E2.1 (clustered retrieval results) (M=6.00, SD=1.10) than using B2.1 (M= 4.88, SD=1.96), although not significantly so, Z=-1.825, p=0.068. It was easier to get started using E2.2 (fielded query) (M= 5.56, SD=1.55) than using B2.2 (M=4.88, SD=1.89), but not significantly so, Z=-0.945, p=0.345.

Subjects felt that it was easier to do searches using E1.1 (database summary) (M=5.13, SD=1.41) than using B1.1 (M=4.81, SD=1.42), but no significant results were found, Z=-0.955, p=0.340. It was significantly easier to do searches using E1.2 (table of contents navigation) (M=5.19, SD=1.22) than using B1.2 (M=3.56, SD=1.67), Z=-2.949, p=0.003. It was significantly easier to do searches using E2.1 (clustered retrieval results) (M= 5.94, SD=1.12) than using B2.1 (M=4.44, SD=1.90), Z=-2.284, p=0.022. It was easier to do searches using E2.2 (fielded query) (M=4.69, SD=2.27) than using B2.2 (M=4.25, SD=1.73), although not significantly so, Z=-0.601, p=0.548.

Subjects were more satisfied with results when using E1.1 (database summary) (M=4.81, SD=1.05) than using B1.1 (M=4.19, SD=1.11), but not significantly so, Z=-1.398, p=0.162. Subjects were significantly more satisfied with results when using E1.2 (table of contents navigation) (M=5.63, SD=0.81) than using B1.2 (M= 4.56, SD= 1.55), Z=-2.738, p=0.006. Subjects were more satisfied with results when using E2.1 (clustered retrieval results) (M=5.19, SD=1.52) than using B2.1 (M= 4.00, SD= 2.00), but not significantly so, Z=-1.283, p=0.199. Subjects were more satisfied with results when using E2.2 (fielded query) (M=5.50, SD= 1.67) than using B2.2 (M=4.06, SD=1.95), but not significantly so, Z=-1.583, p=0.113.

Subjects felt they had more time when using E1.1 (database summary) (M=4.94, SD=1.73) than using B1.1 (M=4.25, SD=1.48), although not significantly so, Z=-1.239, p=0.215. Subjects felt they had significantly more time when using E1.2 (table of contents navigation) (M= 5.75, SD=1.73) than using B1.2 (M=5.00, SD=1.59), Z=-2.080, p=0.038. Subjects felt they had significantly more time when using E2.1 (clustered retrieval results) (M=5.69, SD= 0.95) than using B2.1 (M=4.56, SD=1.79), Z=-2.047, p=0.041. Subjects felt they had more time when using E2.2 (fielded query) (M=6.06, SD=1.24) than using B2.2 (M=5.19, SD=1.42), although not significantly so, Z=-1.528, p=0.126. Figure 5.4 gives a graphical representation of these results.



Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

p < 0.05

Figure 5.4 Statistics of the post-search questionnaire (Experiment I)

Table 5.9

Post-Search Questionnaire Results (Experiment I)

	Systems	Easy getting started?	Easy to search?	Result satisfied?	Enough time?
Mean	B1.1	5.50	4.81	4.19	4.25
	E1.1	5.81	5.13	4.81	4.94
	B1.2	4.00	3.56	4.56	5.00
	E1.2	5.69	5.19	5.63	5.75
	B2.1	4.88	4.44	4.00	4.56
	E2.1	6.00	5.94	5.19	5.69
	B2.2	4.88	4.25	4.06	5.19
	E2.2	5.56	4.69	5.50	6.06
Standard deviation	B1.1	1.03	1.42	1.11	1.48
	E1.1	0.91	1.41	1.05	1.73
	B1.2	1.86	1.67	1.55	1.59
	E1.2	0.70	1.22	0.81	1.73
	B2.1	1.96	1.90	2.00	1.79
	E2.1	1.10	1.12	1.52	0.95
	B2.2	1.89	1.73	1.95	1.42
	E2.2	1.55	2.27	1.67	1.24
p-value	B1.1 / E1.1	0.248	0.340	0.162	0.215
	B1.2 / E1.2	0.002*	0.003*	0.006*	0.038*
	B2.1 / E2.1	0.068	0.022*	0.199	0.041*
	B2.2 / E2.2	0.345	0.548	0.113	0.126

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

*$p < 0.05$

5.2.6 Post-system Questionnaire

Subjects' opinions about the systems were assessed by the post-system questionnaire. Subjects were asked whether the system was easy to learn to use, easy to use, understandable and useful on a 7-point scale, where 1 = "not at all"; 4 =

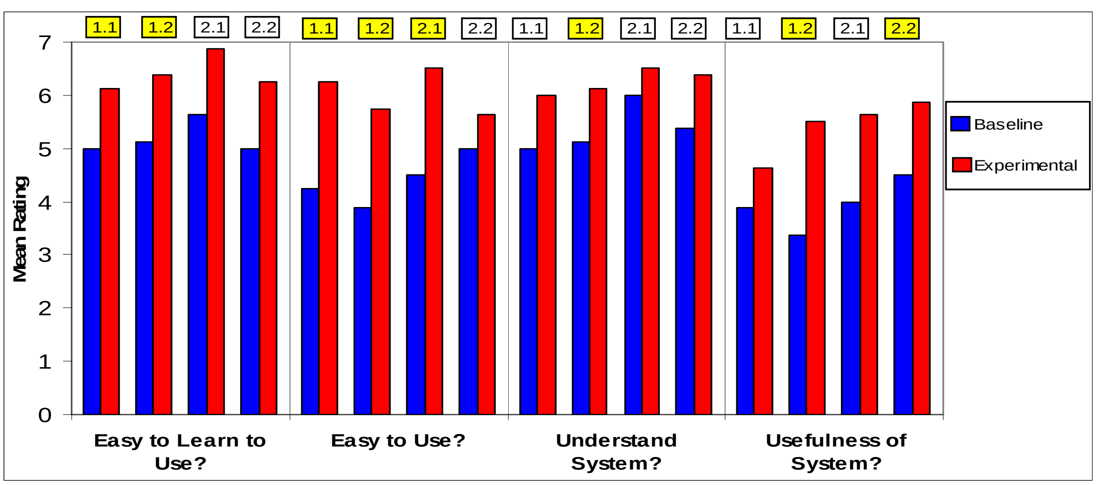
“somewhat”; and 7 = “extremely.” Table 5.10 lists the mean values of each system for these questions. The systems were compared for each question based on each subject’s responses using the non-parametric Wilcoxon signed rank test.

Results showed that it was significantly easier to learn to use E1.1 (database summary) (M= 6.13, SD=0.64) than to use B1.1 (M=5.00, SD= 1.31), Z=-2.060, p=0.039. It was significantly easier to learn to use E1.2 (table of contents navigation) (M= 6.38, SD=0.74) than to use B1.2 (M=5.13, SD= 1.36), Z=-2.060, p=0.039. It was easier to learn to use E2.1 (clustered retrieval results) (M=6.88, SD= 0.35) than to use B2.1 (M=5.63, SD=2.00), although not significantly so, Z=-1.841, p=0.066. It was easier to learn to use E2.2 (fielded query) (M=6.25, SD=1.04) than to use B2.2 (M=5.00, SD=1.69), although not significantly so, Z=-1.276, p=0.202.

Subjects felt it was significantly easier to use E1.1 (database summary) (M=6.25, SD=0.46) than to use B1.1 (M=4.25, SD=1.58), Z=-2.226, p=0.026.It was significantly easier to use E1.2 (table of contents navigation) (M=5.75, SD=0.89) than to use B1.2 (M=3.88, SD=1.36), Z=-2.549, p=0.011. It was significantly easier to use E2.1 (clustered retrieval results) (M=6.50, SD=0.76) than to use B2.1 (M= 4.50, SD=1.77), Z=-2.120, p=0.034. It was easier to use E2.2 (fielded query) (M=5.63, SD=1.30) than to use B2.2 (M= 5.00, SD=1.41), although not significantly so, Z=-0.682, p=0.495.

Subjects felt they understood the system better when using E1.1 (database summary) (M=6.00, SD= 0.76) than B1.1 (M=5.00, SD=1.31), but no significant result was found, Z=-1.511, p=0.131. Subjects felt they understood the system significantly better when using E1.2 (table of contents navigation) (M=6.13, SD=0.64) than using B1.2 (M= 5.13, SD= 1.25), Z=-2.060, p=0.039. Subjects felt they understood the system better when using E2.1 (clustered retrieval results) (M=6.50, SD=0.76) than using B2.1 (M=6.00, SD=1.69), although not significantly so, Z=-0.378, p=0.705. Subjects felt they understood the system better when using E2.2 (fielded query) (M=6.38, SD=0.74) than using B2.2 (M=5.38, SD=1.30), but no significant result to support this, Z=-1.633, p=0.102.

Subjects felt E1.1 (database summary) (M=4.63, SD=1.69) was more useful than B1.1 (M= 3.88, SD=0.99), although not significantly so, Z=-1.163, p=0.245. Subjects felt E1.2 (table of contents navigation) (M= 5.50, SD= 0.76) was significantly more useful than B1.2 (M=3.38, SD=1.30), Z=-2.428, p=0.015. Subjects felt E2.1 (clustered retrieval results) (M=5.63, SD=0.74) was more useful than B2.1 (M=4.00, SD=1.31), although not significantly so, Z=-1.897, p=0.058. Subjects felt E2.2 (fielded query) (M= 5.88, SD=1.13) was significantly more useful than B2.2 (M= 4.50, SD=1.60), Z=-2.060, p=0.039. Figure 5.5 displays the graphical representation of the above results.



Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

□ $p < 0.05$

Figure 5.5 Statistics of the post-system questionnaire (Experiment I)

Table 5.10

Post-System Questionnaire Results (Experiment I)

	System	Easy to learn to use?	Ease to use?	Under-stand system?	Useful-ness of system?
Mean	B1.1	5.00	4.25	5.00	3.88
	E1.1	6.13	6.25	6.00	4.63
	B1.2	5.13	3.88	5.13	3.38
	E1.2	6.38	5.75	6.13	5.50
	B2.1	5.63	4.50	6.00	4.00
	E2.1	6.88	6.50	6.50	5.63
	B2.2	5.00	5.00	5.38	4.50
	E2.2	6.25	5.63	6.38	5.88
Standard deviation	B1.1	1.31	1.58	1.31	0.99
	E1.1	0.64	0.46	0.76	1.69
	B1.2	1.36	1.36	1.25	1.30
	E1.2	0.74	0.89	0.64	0.76
	B2.1	2.00	1.77	1.69	1.31
	E2.1	0.35	0.76	0.76	0.74
	B2.2	1.69	1.41	1.30	1.60
	E2.2	1.04	1.30	0.74	1.13
p-value	B1.1 / E1.1	0.039*	0.026*	0.131	0.245
	B1.2 / E1.2	0.039*	0.011*	0.039*	0.015*
	B2.1 / E2.1	0.066	0.034*	0.705	0.058
	B2.2 / E2.2	0.202	0.495	0.102	0.039*

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

*$p < 0.05$

5.2.7 Exit Questionnaire

Exit questionnaire was presented to the subjects after they completed both systems. Subjects were asked to rate the difference of the two systems on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely."

Subjects were also asked to decide which system was more helpful in completing tasks, was easier to learn to use, was easier to use and which system they liked best, with three choices: system 1 (either E or B), system 2 (either E or B), no difference (ND). In addition, subjects were asked which system features they liked or disliked most and were asked to give some general comments with open-ended questions. For each question, the Sign test (ignoring no difference) was employed to test whether the number of subjects who preferred the experimental systems was significantly different from the number of subjects who preferred the baseline systems.

Subjects found E1.1 (database summary) and B1.1 were different at a high level (M= 5.63, SD=1.06). Subjects found E1.2 (table of contents navigation) and B1.2 were different at a high level (M=6.13, SD=0.35). Subjects found E2.1 (clustered retrieval results) and B2.1 were different at a high level (M=5.88, SD=0.99). Subjects found E2.2 (fielded query) and B2.2 were different at a high level (M=5.75, SD=1.16).

From Table 5.11, subjects found E1.1 (database summary) was more helpful than B1.1, E=6, B=1, ND=1, although not significantly so, p=0.125. Subjects found E1.2 (table of contents navigation) was more helpful than B1.2, E=7, B=1, ND=0, though the difference is not significant, p=0.070. Subjects found E2.1 (clustered retrieval results) was more helpful than B2.1, E=6, B=1, ND=1, but no significant result was found,

p=0.125. Subjects found E2.2 (fielded query) was more helpful than B2.2, E=5, B=2, ND=1, although not significantly so, p=0.453.

Results showed that E1.1 (database summary) was easier to learn to use than B1.1, E=3, B=1, ND=4, but no significant results were found, p=0.625. E1.2 (table of contents navigation) was significantly easier to learn to use than B1.2, E=6, B=0, ND=2, p=0.031. E2.1 (clustered retrieval results) was easier to learn to use than B2.1, E=2, B=0, ND=6, although not significantly so, p=0.500. E2.2 (fielded query) was easier to learn to use than B2.2, E=4, B=1, ND=3, but not significantly so, p=0.375.

Results also showed that E1.1 (database summary) was easier to use than B1.1, E=3, B=1, ND=4, although not significantly so, p=0.625. E1.2 (table of contents navigation) was significantly easier to use than B1.2, E=7, B=0, ND=1, p=0.016. E2.1 (clustered retrieval results) was easier to use than B2.1, E=5, B=1, ND=2, but no significant results were found, p=0.219. E2.2 (fielded query) was easier to use than B2.2, E=4, B=1, ND=3, but not significantly so, p=0.375.

Overall, subjects liked E1.1 (database summary) best, E=5, B=3, ND=0, p=0.727, although the difference was not significant. Subjects liked E1.2 (table of contents navigation) best, E=7, B=1, ND=0, although not significantly so, p=0.070. Subjects liked E2.1 (clustered retrieval results) best, E=7, B=1, ND=0, but no significant result was found, p=0.070. Subjects liked E2.2 (fielded query) best, E=6, B=1, ND=1, p=0.125, but not significantly so. Figure 5.6 shows the graphical representation of the above results.

The results from post-system questionnaires were compared to the exit questionnaires in Table 5.12.

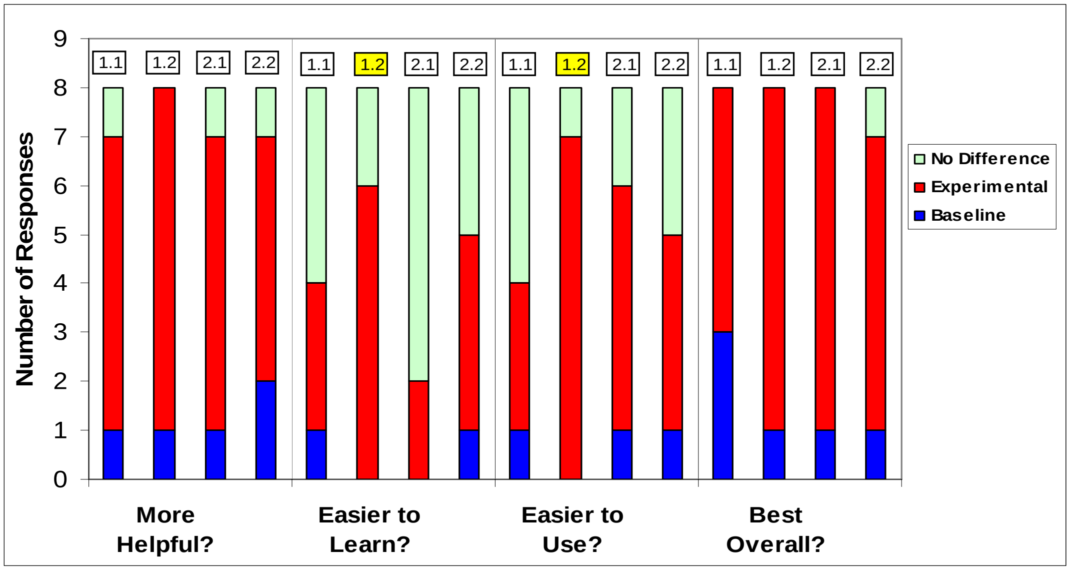
Table 5.11

System Comparison of the Exit Questionnaire (Experiment I)

		Question abstract			
		More helpful?	Easier to learn?	Easier to Use?	Best overall?
System (B1.1/E1.1)	B1.1	1	1	1	3
	E1.1	6	3	3	5
	No difference	1	4	4	0
	p-value	0.125	0.625	0.625	0.727
System (B1.2/E1.2)	B1.2	1	0	0	1
	E1.2	7	6	7	7
	No difference	0	2	1	0
	p-value	0.070	0.031*	0.016*	0.070
System (B2.1/E2.1)	B2.1	1	0	1	1
	E2.1	6	2	5	7
	No difference	1	6	2	0
	p-value	0.125	0.500	0.219	0.070
System (B2.2/E2.2)	B2.2	2	1	1	1
	E2.2	5	4	4	6
	No difference	1	3	3	1
	p-value	0.453	0.375	0.375	0.125

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

*$p < 0.05$



Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

$p < 0.05$

Figure 5.6 Statistics of the exit questionnaire (Experiment I)

Subjects were asked some questions about which system features they liked most and why (see Table 5.13) and what system features they disliked most and why (see Table 5.14). Subjects were also asked to give suggestions and comments about the systems. Examples are provided in Tables 5.13 and 5.14. All quotations are exactly what the subjects wrote, including misspellings and other errors.

Briefly speaking, subjects liked table of contents navigation because of the "hierarchical structure" and the way that more context can be seen. Subjects liked the ranked list of paragraphs because of the "simplicity". Subjects liked database summary because of the overview of the databases. Subjects liked the ranked list display of the database results because of the "integrated list" and the display of titles. Subjects liked clustered retrieval results because of the clusters and because it was "easy to use". Subjects liked the ranked list of documents because of the simplicity and neatness.

Subjects liked the feature of fielded query because more options were given and the convenience of displaying all information. Subjects liked the feature of a ranked list of complete citations of books because it allowed them to “get to table of contents” from the citation.

Generally, subjects didn’t like the ranked list of documents with indication of which database this document was in because it lacked an “overview of results in each database” and it was unorganized, so they could not get a sense about the relationship between the documents in the databases. Subjects didn’t like database summary because it was not very helpful for the specific task. Subjects didn’t like table of contents because it allowed people to see the chapter title on the same window. Subjects didn’t like the ranked list of paragraphs because there was “no (apparent) way to search for text”. Subjects didn’t like clustered retrieval results because some of the labeled “words under cluster were not helpful”. Subjects didn’t like the ranked list of documents because the documents were not “categorized”. Subjects didn’t like fielded query because some results were not very useful. Subjects didn’t like the ranked list of complete citations of books because it was easy to get confused.

In the end, subjects gave many suggestions and comments about the systems. Generally, regarding database summary, subjects said they would like to have “a short information of the article along with title“; “more instructions about what can/cannot do in the use of the system would be more helpful”; “work on eliminating the system delays in system 1 (returning from doc to list) or, at least change the mouse cursor to an hourglass"; "redesign the rankings box to make the descriptive text permanent and use rank change buttons for the databases (more up/down the list)”; “how about to rank the

DB by the correlation of the keywords. For example, rank them by the highest correlation rate of the number of keywords"; "When I click the article title, some results are full text of the article, some results are abstract. I wish every title can link to its abstract. It will be easy to tell if the article is related to the topic."

Regarding table of contents navigation, subjects said they would like to "highlight the name or place of preference. Ex: The individual name and highlight the quotation"; "Use color to identify key point (but it will no use for blind color people though)"; "the text should fit entirely in the text box-scrolling shouldn't be required."

Regarding clustered retrieval results, subjects said they would like to "1) add which links (both clusters and docs) have been touched, 2) keep position in the list (both clusters and ranked list) when a document is examined"; "it is better if I can search the keyword inside each article, and it is better if the system keep track of what I have read"; "included copy and paste button in the search box or add these functions to right click menu of mouse."

Regarding fielded query, subjects said they would like to "maybe put the author's gender as a limitation item for search in case sb. only remembers the gender of the author"; "add like 'keyword' field to reach more detailed info in database."

When subjects talked about which features they preferred, they always referred to features used in Google. For example, one subjects said the reason that he liked the fielded query search was "because it just seems that you are doing advanced search using Google." In entry questionnaire, subjects were asked to describe their favorite search engine. All subjects wrote down "Google."

Table 5.12

Comparison of the Post-system and the Exit Questionnaire (Experiment I)

Questionnaires	Systems	Easier to learn?	Easier to use?
Post-system	B1.1	0.039*	0.026*
	E1.1		
	B1.2	0.039*	0.011*
	E1.2		
	B2.1	0.066	0.034*
	E2.1		
	B2.2	0.202	0.495
	E2.2		
Exit	B1.1	0.625	0.625
	E1.1		
	B1.2	0.031*	0.016*
	E1.2		
	B2.1	0.500	0.219
	E2.1		
	B2.2	0.375	0.375
	E2.2		

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

*$p < 0.05$

Table 5.13

IR Support Techniques that Subjects Liked (Experiment I)

IR support techniques	Reasons that subjects liked the feature
Database summary	“I liked the general overview of the results for all database, even though I did not trust those numbers to necessarily indicate relevance.”; “drilling down database to see related documents”; “the database grouping were most helpful”
Ranked list of documents with source of each document	“the integrated list was very helpful.” ; “display of article titles, even though not all of them are very informative”
Table of contents navigation	“hierarchical structure”; “the folder layout”; “navigation menu in the left”; “more user friendly”; “TOC gives the structure of the book”; “ability to click on chapter to see sections; ability to pull up the text in the section by clicking on the section.”; “gives more context”
Ranked list of paragraphs	“text fragments for browsing”; “simplicity”
Clustered retrieval results	“clusters - so I could see list of choices on one screen because info is provided in manageable chunks; I notice I am drawn to items that are capitalized believeing that font signals relevance”; “copy/pase search results/ clustering documents”; “shorter search space for each cluster”; “cluster keywords”; “terms in cluster;” “clustered information, the articles that are most related to the topic were shown under cluster, easy to use”
Ranked list of documents	“the interface is quite simple and neat”
Fielded query	“I had several options for searching publish year, range of publication year or exact year”; “boolean syntax; keyword indexing, query "airplance" text "areoplane"”; “field search, display TOC”; “publication place and publication year” “year and year range”; “convenient to show you all the information”
Ranked list of complete citations of books	“can get to table of content”

Table 5.14

IR Support Techniques that Subjects didn't Like (Experiment I)

IR support techniques	Reasons that subjects didn't like the feature
Database summary	"The display of the keywords and their potentially relevant documents. I did not know the search results are displayed, and the relationships between each keyword and overall search results. It might be helpful for choosing a database to search, but I did not find it potentially useful in these tasks."; "separate documents by sources"
Ranked list of documents with source of each document	"No overview of the results from each database"; "one needs to organize in a way so that users can get ideas whether they will use the specific one or not."; "The results in the same databases. Don't arrange together."
Table of contents navigation	"allows you to see the chapter title on the dame window"
Ranked list of paragraphs	"no right click to cut and paste"; "no (apparent) way to search for text strings; it wasn't clear what it was offering"
Clustered retrieval results	"the words under cluster were not helpful in defining if this very cluster is the one you need to choose to get access to the necessary infor"
Ranked list of documents	"I didn't like the fact that I cannot change search items; it would be nice if search terms are highlighted in documents"; "ranked list length"; "not categorized. The documents seem not ranked from the top to the bottom"
Fielded query	"there are some useless results, ie. The combination of one search fileld without the others"; "did not have time to experience limits of boolean (e..g phrase "new york" wildcard indian nesting etc.)"; "the field of title has to be exact match, could not get to "table of content""
Ranked list of complete citation of books	"easy to let you confuse your search results"

Chapter 6

Research Problem 1: Discussion

This chapter discusses the results of the first experiment from three perspectives: performance, interaction, and usability. Considering all three criteria, there is reasonable support for the general hypothesis that the systems tailored to support specific ISSs were “better” than generic baseline systems using techniques to support specified search.

6.1 Performance

The performance results supported two hypotheses of this study. The system with table of contents navigation performed better in supporting scanning tasks than the baseline system, and the system with fielded query performed better in supporting searching tasks than the baseline system.

Hypothesis 1, that the database summary technique can improve system performance in supporting scanning tasks was not significantly supported by the results. Subjects using E1.1 (database summary) felt more satisfied with the results and spent less time than those of B1.1, but not significantly so.

Hypothesis 2, that the table of contents navigation technique can improve system performance in supporting scanning tasks was supported by the results. Subjects using E1.2 (table of contents navigation) spent significantly less time and felt significantly more satisfied with the results than those of B1.2. Otherwise, no significant results were found.

Hypothesis 3, that the clustered retrieval results technique can improve system performance in supporting searching tasks was not significantly supported by the results. Subjects using E2.1 (clustered retrieval results) felt more satisfied with the results, spent less time, and identified more relevant aspects than those of E2.1, but not significantly so.

Hypothesis 4, that the fielded query technique can improve system performance in supporting searching tasks was supported by the results. Subjects using E2.2 (fielded query) spent significantly less time than those using B2.2. This indicated that fielded query was a technique that saved effort by reducing time. Otherwise, no significant differences were found.

Subjects spent less time using the experimental systems than using the respective baseline systems, although differences were not significant in two of the four cases. It was found that subjects using system E1.1 (database summary)/B1.1 and E2.1 (clustered retrieval results)/B2.1 spent more time than using system E1.2 (table of contents navigation)/B1.2 and E2.2 (fielded query)/B2.2. In fact, many subjects used up the maximum time of ten minutes in doing searches using E1.1/B1.1 and E2.1/B2.1. This could be attributed to two reasons. One is that the tasks performed in these two sets of systems might be more difficult than the same tasks performed in the other two sets of systems. The other is that the subjects were asked to identify the best databases (for E1.1/B1.1) or as many aspects of the given topics as possible (for E2.1/B2.1), which could be naturally more time consuming. The results indicated that these tasks motivated subjects to spend time to explore as much as they can. They also indicated that subjects were not sure whether they found out all that they should find out from the system, or they were not sure whether what they found covered all aspects of the given topic.

For systems E1.2 (table of contents navigation)/B1.2 and E2.2 (fielded query) /B2.2, subjects using the experimental system got more correct answers than using the baseline system. This indicated that the table of contents navigation technique and fielded query technique can help subjects get effective answers. The reasons underlying this could be that the table of contents navigation provided subjects with more context, while the fielded query technique offered subjects more choices which were appropriate to the exact task. For example, system E2.2 displays fields of a book such as author, title, publication year, place and publisher. This could also explain why subjects using these two systems spent less time to get results. However, since no significant difference was found to confirm this result, it is not safe to conclude this way.

Subjects using E2.1 (clustered retrieval results) identified more relevant aspects of the topic than those using B2.1. This result indicated that the technique of grouping documents into clusters is possibly an effective way to encourage better searches, but because the difference was not significant, this can only be an indicative result.

6.2 Interaction

Subjects using the system with fielded query (E2.2) had fewer iterations and got more correct answers than those using the baseline system (B2.2). Although this result was not significantly supported, it is still an indication that fielded query helped subjects get more effective answers by less interaction with the system.

Subjects using the system with clustered retrieval results (E2.1) ultimately saved fewer documents and identified more relevant aspects than those using the baseline system (B2.1). Although the differences were not significant, it indicated that clustered retrieval results could help subjects get more effective answers by saving fewer documents.

Subjects using the system with database summary (E1.1) viewed more documents than those using the baseline system (B1.1). Subjects using the systems with clustered retrieval results (E2.1) and fielded query (E2.2) viewed fewer documents/books than those using the respective baseline systems. Considering that the subjects using experimental systems also identified more relevant aspects or more correct answers, these are good indications that clustered retrieval results and fielded query helped subjects get more effective answers with less interaction with the system and less effort.

6.3 Usability

Results from the post-system questionnaires strongly demonstrated that the systems which incorporated the techniques tailored to different ISSs were more usable than the baseline systems.

Specifically, the system with database summary (E1.1) was significantly more usable than the baseline system (B1.1) in terms of ease of learning to use, and ease of use. The subjects found that the experimental system was more usable than the baseline system with respect to understandability and usefulness, although the differences were not significant. It could be attributed to the reason that all subjects claimed that they liked Google which uses ranked lists of documents. Intuitively, the baseline system should be more or equally easy to learn and use than the experimental system. But this study got the opposite results. This could be reasonably explained by the findings from exit questionnaires, that the overview of the databases makes the system with database summary (E1.1) easier to learn and easier to use.

The system with table of contents navigation (E1.2) was significantly more usable than the baseline system (B1.2) with respect to ease of learning to use, ease of use, understandability, and usefulness. This result indicated that for a task about finding comments from books, table of contents navigation was a better IR support technique than a ranked list of paragraphs because table of contents navigation provided a hierarchical structure of the book and gave more contextual information (e.g., chapter titles).

The system with clustered retrieval results (E2.1) was significantly more usable than the baseline system (B2.1) in terms of ease of use. It was also confirmed that clustered retrieval results was a better technique than the ranked list of documents for this specific ISS and task because clustered retrieval results categorized the documents into clusters and labels each cluster with frequently appearing terms in each cluster. E2.1 was more usable than B2.1 with respect to understandability, although not significantly so. The failure to achieve significance could be because most subjects were already used to IR systems (such as Google) which provided a ranked list of documents as the display results.

The system with fielded query (E2.2) was significantly more usable than the baseline system (B2.2) in terms of usefulness. E2.2 was more usable than B2.2 with respect to ease of learning to use, and ease of use, but the differences were not significant. Intuitively, fielded query was easy to learn to use because it was quite similar to the advanced features provided by popular search engines such as Google. But it was not easy to use because it required the subject to be familiar with some intrinsic rules about Boolean search and fielded search. By tailoring to a specific task, the complexity and difficulty were increased. The results also indicated that fielded query was a better technique than a generic query because it provided more choices for the subject to type in the query in different fields (e.g., author, publication year). Thus any clue the subject had about the book were able to be incorporated in the search and helped the subject get the final result.

The above results suggested that a technique tailored to a specific strategy and task by providing more context or more choices, was more usable than the generic one, that is, a ranked list of documents or paragraphs/complete citations of books.

The results from exit questionnaires showed that the system with table of contents navigation (E1.2) was significantly easier to learn to use and easier to use than the baseline system (B1.2), which were consistent with the findings from post-system questionnaires. But the results from the post-system questionnaires that the system with database summary (E1.1) was significantly easier to learn to use and easier to use than the baseline system (B1.1) have not been confirmed with the exit questionnaires. This could be attributed to the inconsistent ratings assigned by the subjects.

Although the results from the exit questionnaires showed that more subjects thought the experimental systems were helpful than the baseline systems, the difference was not significant. However, it indicated that systems employing task-specific IR support techniques could better support subjects in doing some specific tasks than the equivalent baseline systems.

The measures on which the experimental systems significantly outperformed the baseline system are summarized in Table 6.1. There were no cases in which the baseline system significantly outperformed the experimental system.

Table 6.1

Measures with Significant Results Favoring the Experimental System (Experiment I)

Categories		Measures	Systems	p-value
Performance		Time	E1.2	0.009
			E2.2	0.003
		Result satisfied	E1.2	0.006
Usability	Post-search questionnaire	Easy getting started	E1.2	0.002
		Easy to search	E1.2	0.003
			E2.1	0.022
		Enough time	E1.2	0.038
			E2.1	0.041
	Post-system questionnaire	Easy to learn to use	E1.1	0.039
			E1.2	0.039
		Easy to use	E1.1	0.026
			E1.2	0.011
			E2.1	0.034
		Understand system	E1.2	0.039
		Usefulness of system	E1.2	0.015
			E2.2	0.039
	Exit questionnaire	Easier to learn	E1.2	0.031
		Easier to use	E1.2	0.016

Note. E1.1: database summary; E1.2: table of contents navigation; E2.1: clustered retrieval results; E2.2: fielded query.

Chapter 7

Research Problem 1: Conclusions

It can be concluded that systems with specific IR support techniques can better support different ISSs for different tasks than the respective baseline systems.

Hypothesis 1 (database summary) was not confirmed with respect to performance. But the questionnaire results showed that the system with database summary was significantly more usable than the system with a ranked list of documents in terms of ease of learning to use and ease of use. Thus, despite the relatively small number of subjects, it seems safe to conclude that the database summary system, designed explicitly to support the task of finding the best databases, was better than the generic system with a ranked list of documents.

Hypothesis 2 (table of contents navigation) was confirmed, when performance was measured as time and result satisfaction. Since there was no difference in the accuracy or completeness of an answer, decreased effort and increased satisfaction indicated that table of contents navigation was an effective technique in helping improve search performance in this specific scanning task. The usability results showed that the system with table of contents navigation was significantly more usable than the system with a ranked list of paragraphs in terms of ease of learning to use, ease of use, understandability, and usefulness. Thus, despite the relatively small number of subjects, it seems safe to conclude that the table of contents navigation system, designed explicitly to support the task of finding comments from books, was better than the generic system with a ranked list of paragraphs.

Hypothesis 3 (clustered retrieval results display) was not confirmed with respect to performance. But the questionnaire results showed that the system with clustered retrieval results was significantly more usable than the system with a ranked list of documents in terms of ease of use. Thus, despite the relatively small number of subjects, it seems safe to conclude that the system with clustered retrieval results display, designed explicitly to support the task of finding relevant documents to a given topic, was better than the generic system with a ranked list of documents.

Hypothesis 4 (fielded query) was confirmed, when performance was measured as time. Results on usability also showed that the system with fielded query was more usable than the system with generic search in terms of usefulness. Thus, despite the relatively small number of subjects, it seems safe to conclude that the fielded query system, designed explicitly to support the task of finding a (partially) known item, was better than the generic unstructured query system intended to support search in general.

These results indicated that different ISSs can be better supported by different IR support techniques for different kinds of tasks. Our results also supported that the systems which incorporated those techniques tailored to different ISSs were more usable than the baseline systems. These results indicated that an IR system should be designed by incorporating different IR support techniques for different ISSs and tasks.

Results from exit questionnaire data showed that subjects took great advantage of contextual information (for example, the structure of the table of contents in the navigation interface, and different fields in the fielded query interface) when searching for information. This indicated that an IR system should be designed to provide more contextual information. It also indicated that it is important to conduct research in *contextual IR* in order to find out more about how contextual information help subjects in their searching for information and to investigate the relationships between varieties of contextual factors, such as tasks, stages and goals.

In research problem 3, the foregoing results informed the design of the integrated system which allows seamless change from one ISS to another by incorporating the dialog structure used by MERIT (Belkin et al., 1995). To implement such a system, designing/adopting dialogue structures (research problem 2) which can control the sequence of interactions between the subject and the system is essential.

The generalizability of this study is limited in several aspects.

- Sample size and topic size. Each set of experiments only had 8 subjects and each subject searched only four tasks. The small sample size and topic size limited the generalizability of the results.
- Number of IR support techniques. Four IR support techniques — database summary, table of contents navigation, clustered retrieval results and fielded query — were tested in four experiments respectively. Four techniques are a small portion of the large variety of existing IR support techniques. More IR support techniques need to be explored. Therefore, this study could only conclude that all these four techniques were good candidates but how they compare to the entire universe of IR support techniques still awaits further investigation.
- Types of ISSs, tasks and topics. Based on the classification scheme of Cool and Belkin (2002), there are multiple ISSs. For each ISS/task, the coverage of topics could be widely distributed. In this study, only two types of ISS/task with four

topics each were investigated. Future research should be done using more tasks for more specific categories of ISSs.

- Characteristics of the subject sample. Most of the subjects majored in library and information studies. Some of them had a master's degree in information studies already. This may bias the results because they had more experience in searching. As such, the results of this study might not be easily generalized for novice or less experienced searchers.

Chapter 8

Research Problem 2: Dialogue Structure and System Design

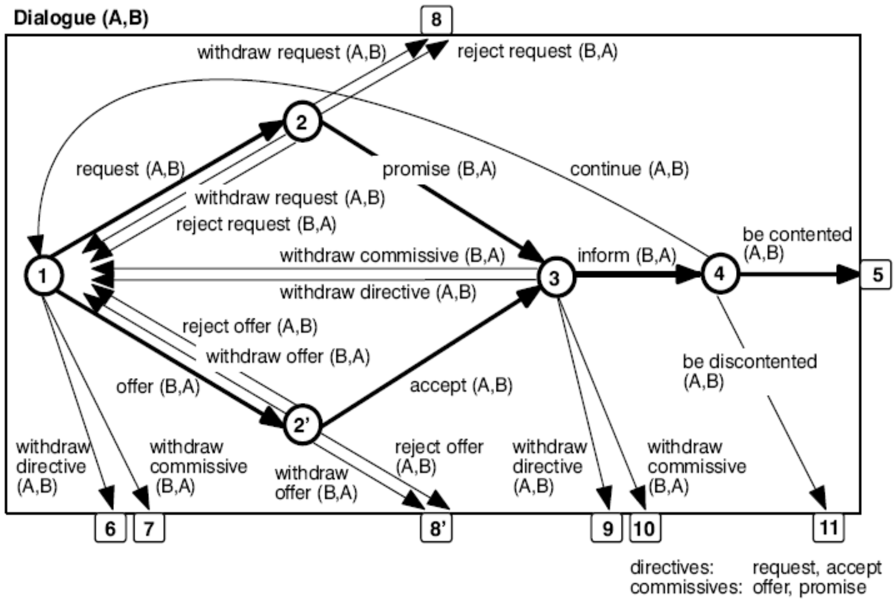
This chapter describes a dialogue structure incorporated in the experimental (integrated) system in experiment II. It starts with the formal information-seeking dialogue model, which is the "Conversational Roles" (COR) model (Sitter & Stein, 1992) with script-based user guidance (Belkin et al., 1995). Then it follows with the detailed design of the dialogue structures and scripts in this study. It also discusses the design and implementation details of the experimental and baseline systems.

8.1 Specifying a Dialogue Structure for Information-Seeking

The interaction of IR occurs when the user interacts with the system, and with the information objects. Interaction with the system can be viewed as a dialogue between the user and the computer through the interface for the purpose of effectively using information by affecting the user's cognitive state (Saracevic, 1997; Belkin et al. 1995; Ingwersen, 1992). In other words, human-computer dialogues/interactions work like equivalent conversations among human beings. During the conversation, one user initiates some issues to discuss, and others respond and may come out with more issues as the conversation continues. However, interactions are much more complicated than conversations. Basically, a user conducts a dialogue through an interface by making utterances (e.g., commands) and receiving responses from the computer. To a large extent, whether an interaction is effective is attributed to the design of an effective dialogue structure. Therefore, the dialogue structure should consider not only the illocutionary aspect, but also the communicative effects of dialogues (Reichman, 1985).

The experimental system proposed in this study supports several different ISSs, which are either scanning-based or searching-based. In order to guide the presentation of specific IR support techniques during the course of an information-seeking episode, it is necessary to specify a dialogue structure. This dialogue structure is equivalent to a dialogue manager, and can be used to control the interactions between the user and the system. This study employed the pre-existing dialogue structure model developed in the MERIT system (Belkin et al., 1995), which fits this study well because it models human-computer interaction as dialogues, and because particular dialogue structures are associated with different ISSs in this model. This model was incorporated into the experimental system at the user interface level and acted as the dialogue manager.

The MERIT system employs the COR model and a script-based user guidance to direct human-computer interaction. The COR model is a model of information-seeking dialogue which describes the interaction between the user and the system at the discourse act level. The underlying assumption of the COR model is that the user and the system act cooperatively. The COR model defines a variety of dialogue acts and local patterns of information exchange between the user and the system. Figure 8.1 gives some idealized courses of a dialogue (using bold arrows between states <1> to <5>). For example, A starts a dialogue by requesting information, B promises to take care of it and presents the information to A, A is satisfied with the given information and then terminates the dialogue. However, information-seeking dialogues are usually highly structured and much more complicated. For instance, A begins with a request "search for a digital camcorder", then B asks some additional information "Would you like to see an overview first?" In such a case, a sub-dialogue will be conducted to clarify this request.



A: Information seeker

B: Information provider

(A,B): Speaker(A)-hearer(B)

→ : Transitions between two states

Figure 8.1 Basic COR dialogue schema (after Belkin et al., 1995)

Each ISS could be associated with an interaction pattern of moves, which can be construed as a script of a dialogue between the user and the system. Belkin et al. (1995) proposed that each ISS is related to a set of moves. They further suggested that each ISS can be connected with a hierarchy of goals or a goal tree which lays out the related moves. These are the basis of the MERIT system (Belkin et al., 1995). As soon as a task has been selected, the MERIT system accesses its saved plan of moves by identifying a script that was associated with this task. Scripts are dialogues and idealized discourses related to dialogue participants and the tasks. A dialogue manager is responsible for tracking the interaction between the user and the system, as well as determining the next dialogue in terms of the user's behavior.

The experimental system — Multiple Information-Seeking Strategies (MISS) – designed in this study adopted the dialogue structure incorporated in MERIT system. The following examples show the dialogue structures associated with different ISSs.

8.2 Standard Introduction Session

An interaction starts with a standard introduction session (see Table 8.1). This session summarizes the functionalities of the system by telling the user which kinds of interaction it can support. Once the user chooses one interaction type, the system shows the user the process of such an interaction. After the user has learned about the process, the specific interaction starts.

Table 8.1

Dialogue Structure of the Introduction Session

1. system:	Here's a list of choices.
2. user:	I am interested in this (chooses one).
3. system:	O.K. This is what I can do for you. → 4 or 1 or 5
4. user:	a. Let's do it. → 5
	b. No. I don't like this. → 1

At the discourse act level, this script can be seen as a complete dialogue cycle based on the COR model, that is, A: offer → B: accept → A: inform → B: continue.

8.3 Example Dialogue Structures for Searching/Scanning

When ISSs are described as a sequence of interactions between the user and the IR system, dialogue structures play the role of identifying patterns of such interactions. In each dialogue structure, the system predicts the next move of the user and provides alternative options for the user to proceed.

In the situation when a user has a vague recollection about a book that s/he saw, s/he needs to improve her/his knowledge of some characteristics of the book, such as topic, author, and title. Thus, s/he might need to search the system on terminological fragments of those data elements. Another situation is that a user is interested in one particular topic and wants to find more documents of this topic, and s/he needs to construct a systematic search within one database for the particular topic in order to find documents of interest.

Table 8.2 gives a dialogue structure of a searching strategy.

Table 8.2

Dialogue Structure I: for Searching

5. user:	I want to find something that corresponds to this (specification of kinds of items to be retrieved);
6. system:	a. Here is what I find.
	b. I can't find anything like what you asked for. → 7b
7. user:	a. 1. I want to continue. → 7b
	2. I want to quit. →10
	3. I want to look at this in more details (selects one from list). → 7a.4 or Dialogue Structure II
system:	4. Here are the details.
user:	5. I like this one. →8
	6. I don't like this. →7a.8
	7. Show me more details. →7a.4
system:	8. How about one of these (shows the list)? →7a.1, 7a.2 or 7a.3, or if nothing left in the list →7b or 8
system:	b. Would you like to try other ways to find what you asked for?
user:	1. O.K. Let's do it. → 5 or 1 or Dialogue Structure II
	2. No. I want to quit. →10
8. system:	Shall we save this and continue? →9
9. user:	a. Yes. →7a.8 or 7b
	b. No, just continue. →7a.8 or 7b
	c. No, just quit. →10
	d. No, save this then quit. →10
10. system:	Goodbye.

At the discourse act level, this dialogue structure can be described as follows according to the COR model.

1. Steps 5-7a are a dialogue cycle for the purpose of specifying items to be retrieved from the database.

2. Step 7b offers a tactic for continuing a new dialogue.

3. Steps 8-9 are initiated by the system to request advice from the user on how to proceed.

Interactions in this dialogue structure demonstrate an example of a searching strategy defined by interaction by searching with the method of search and the mode of specification in an information object. By employing such an interaction sequence, the user would be able to specify a search criteria based on the retrieval results, as well as modify the initial specification based on a set of ISSs relevant to a particular situation.

Considering a situation that a user is interested in one particular topic but has no idea about which of many possible databases to search, s/he needs to identify the best databases for the particular topic: that is, rank them. Another situation is that a user is preparing a talk for a conference. S/he recalls some germane comments from a known electronic book but cannot remember the exact contents. S/he needs to find out the exact quotations. To accomplish these two tasks, the user needs to use a scanning strategy. Table 8.3 gives a dialogue structure of a scanning strategy.

Table 8.3

Dialogue Structure II: for Scanning

5. system:	a. Here is the overall structure of the system. → 8 b. You can select a starting point from which to view the structure. → 6
6. user:	I want to start from this one (selects a starting point) → 7
7. system:	Here is the structure/contents. → 8
8. user:	a. Show me the structure/contents associated with this item (selects from display). → 7 b. I want to use a different starting point → 5b c. I want to search the contents in the structure. → Dialogue Structure I d. I like this one. → 9 e. I don't like this one, but want to continue. → 7 or Dialogue Structure I f. I want to stop this and do something else. → 1 g. I want to quit. → 11
9. system:	Shall we save this and continue? → 10
10. user:	a. Yes. → 8a or 8b b. No, just continue. → 8a or 8b c. No, just quit. → 11 d. No, save this then quit. → 11 e. Save this and do something else. → 1 f. Don't save this and do something else. → 1
11. system: Goodbye.	

At the discourse act level, this dialogue structure can be described as following according to the COR model.

1. Steps 5-8 are a dialogue cycle for the purpose of recognizing interesting items through scanning the system.

2. Steps 9-10 are initiated by the system to request advice from the user on how to proceed.

Interactions in this dialogue structure demonstrate an example of a scanning strategy defined by interaction by scanning with the method of scan and the mode of recognition in a meta-information resource.

8.4 Implementing and Evaluating an Experimental System Supporting Multiple ISSs

This system incorporates the four techniques that proved effective in the first set of experiments, as well as the dialogue structures proposed in the preceding sections.

8.4.1 General Design Issues

Both the baseline system and the experimental system were constructed using Java and the LEMUR Toolkit, using Indri indexing, passage indexing, structured-query retrieval and document clustering. As test collections, we used the 2004 TREC HARD collection of eight news databases, and a specially prepared database of 50 books downloaded from Project Gutenberg. Details of the system and the databases are provided in section 4.2 of Chapter 4.

Both the baseline and experimental system have the same general interface structure. They begin with an introductory screen, asking the user to choose one of several functionalities. Choosing one leads to a screen which has a query box and "search" button at the top, a large results display area, and a column on the right, the top of which displays the topic and the bottom a space for saving results, with a horizontal bar across the bottom of the screen with navigation buttons.

8.4.2 Experimental System Design

The experimental system allows the user to use a variety of ISSs and to seamlessly switch from one ISS to another in a single information-seeking episode. The system suggests appropriate ISSs to the user at the appropriate time, given the current state of the information-seeking episode. The following flow chart (see Figure 8.2) describes how this system is constructed.

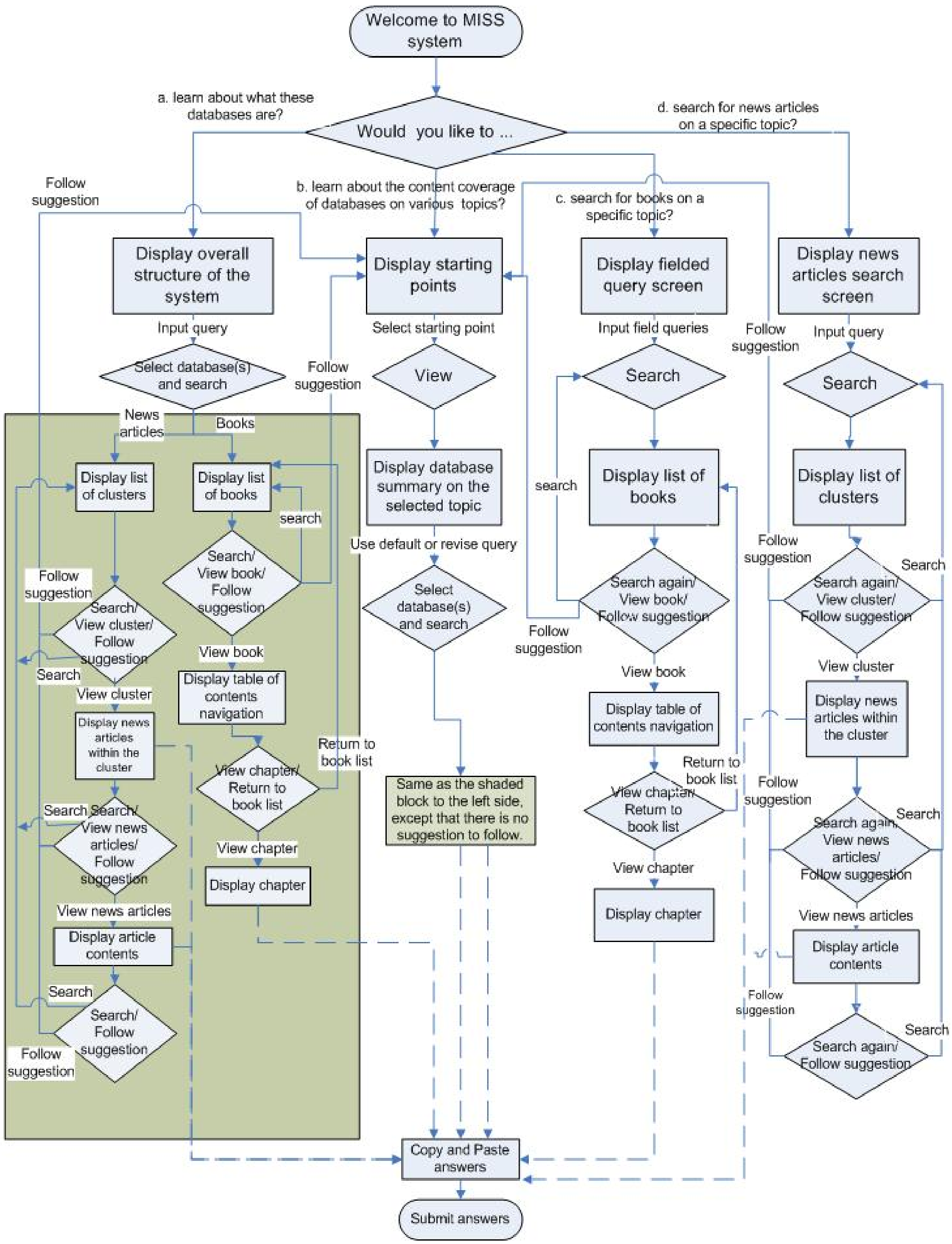


Figure 8.2 Flow chart of the experimental system

8.4.3 Experimental System Implementation

The experimental system incorporated the dialogue structures discussed above, and the IR support techniques used in the experimental systems in research problem 1. These IR support techniques include database summary, clustered retrieval results, table of contents navigation and fielded query.

8.4.3.1 Welcome Screen

The system starts with a welcome screen: "Welcome to MISS system. There are nine databases in the system." It provides four options: learning about what the databases are; learning about content coverage of databases on various topics; searching for books on a specific topic; and searching for news articles on a specific topic (see Figure 8.3). By choosing different options, the user can search for information on a variety of topics. Eight databases contain news articles, and one has books.

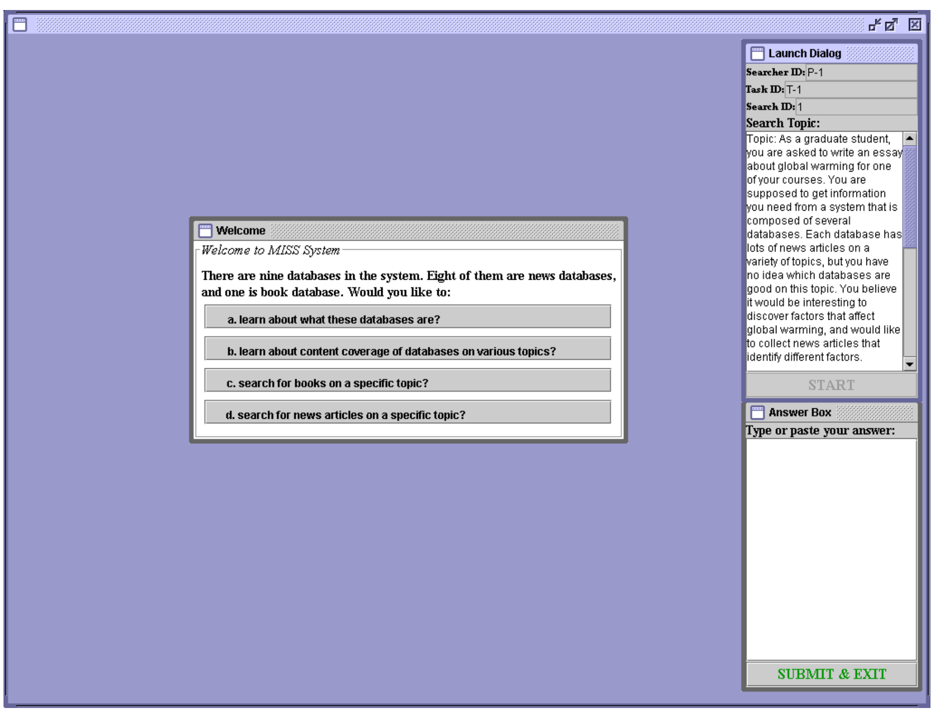


Figure 8.3 Welcome screen of the experimental system

8.4.3.2 Learn about the Overall Structure of the System

When the user chooses option "a. learn about what these databases are," the system provides the name of each database with a description about the number of documents in each. A search box for query input and a search button for initiating the search are given on the top of the screen. Thus the user can formulate or reformulate a query and search on one or multiple databases.

If the task is about finding documents, after the query is typed in and the database(s) are selected (see Figure 8.4), the clustered retrieval results are displayed (see Figure 8.5). After clicking on a cluster, all the related documents in that cluster are shown (see Figure 8.6). The system sets a threshold of 30 as the maximum number of documents in each cluster. Also it provides the source of each document. Clicking on the title of each document leads to display of the document content (see Figure 8.7).

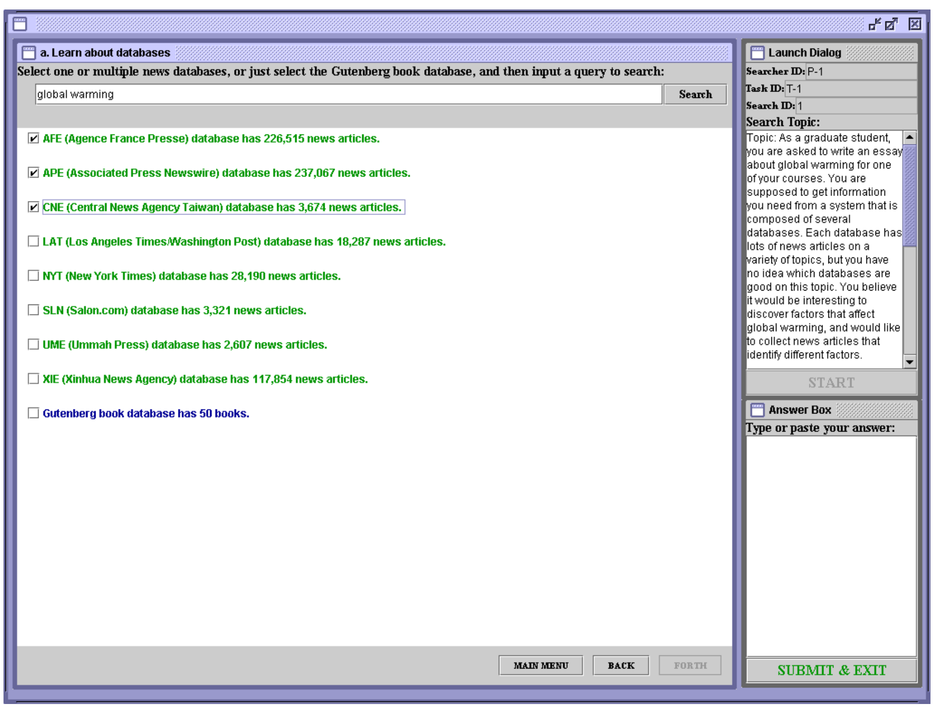


Figure 8.4 Overall structure of the experimental system (I)

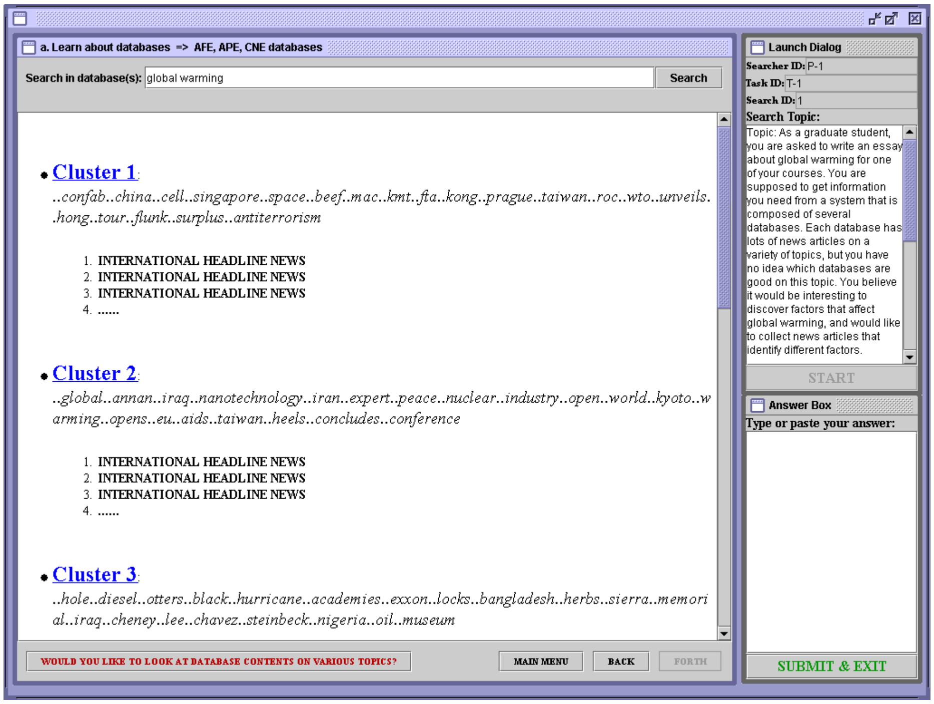


Figure 8.5 Clustered retrieval results of documents

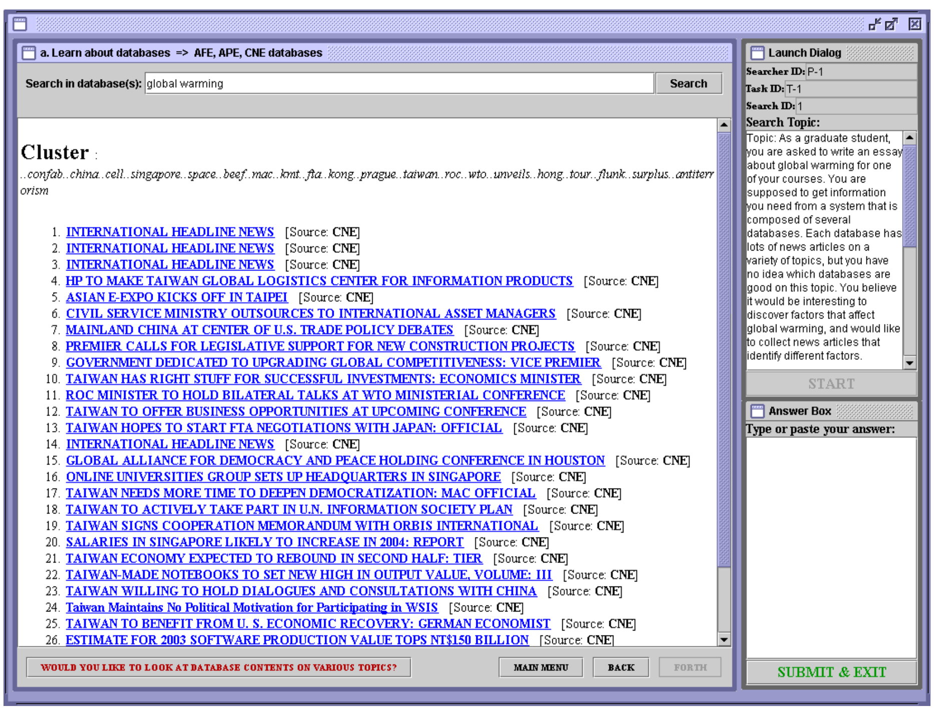


Figure 8.6 Contents of the cluster

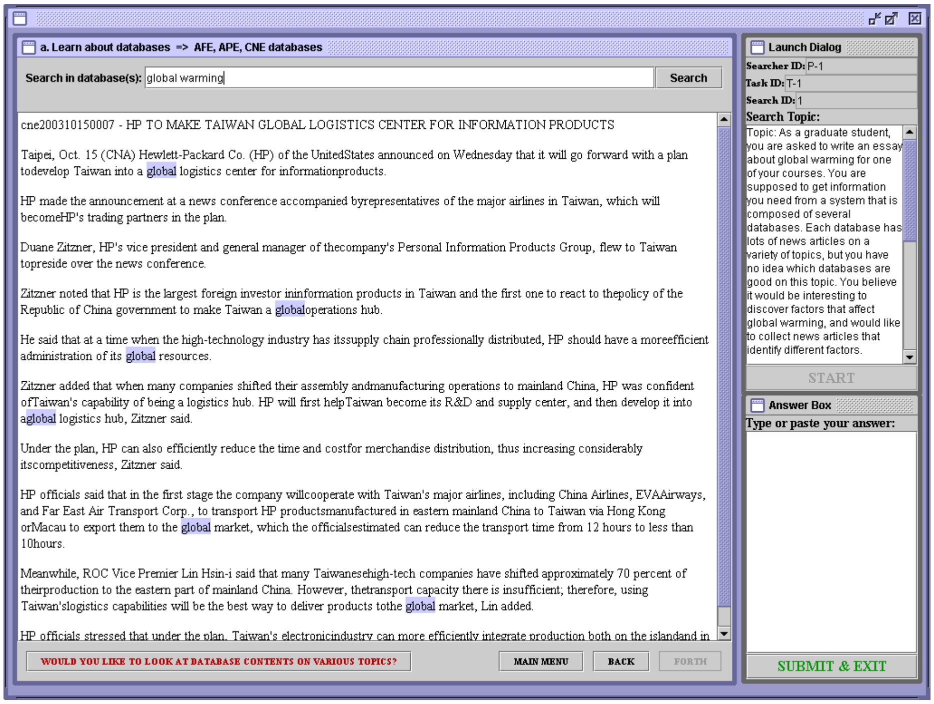


Figure 8.7 Contents of the document

If the task is about finding books, after the query is typed in and the database(s) are selected (see Figure 8.8), a ranked list of complete citations of books is displayed (see Figure 8.9). After clicking on each book, the table of contents of that book is shown (see Figure 8.10).

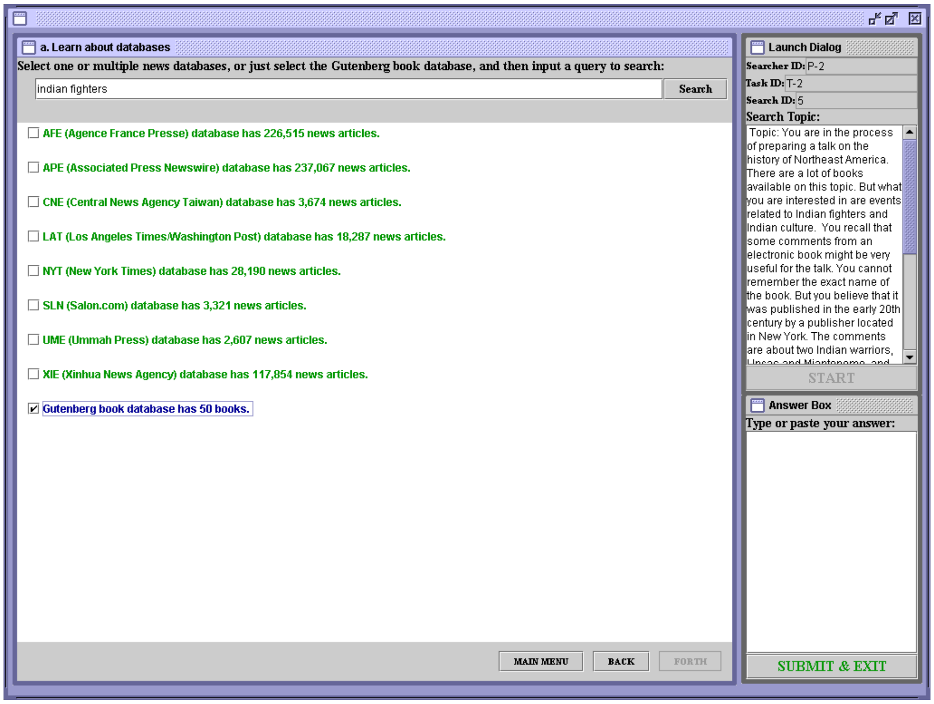


Figure 8.8 Overall structure of the experimental system (II)

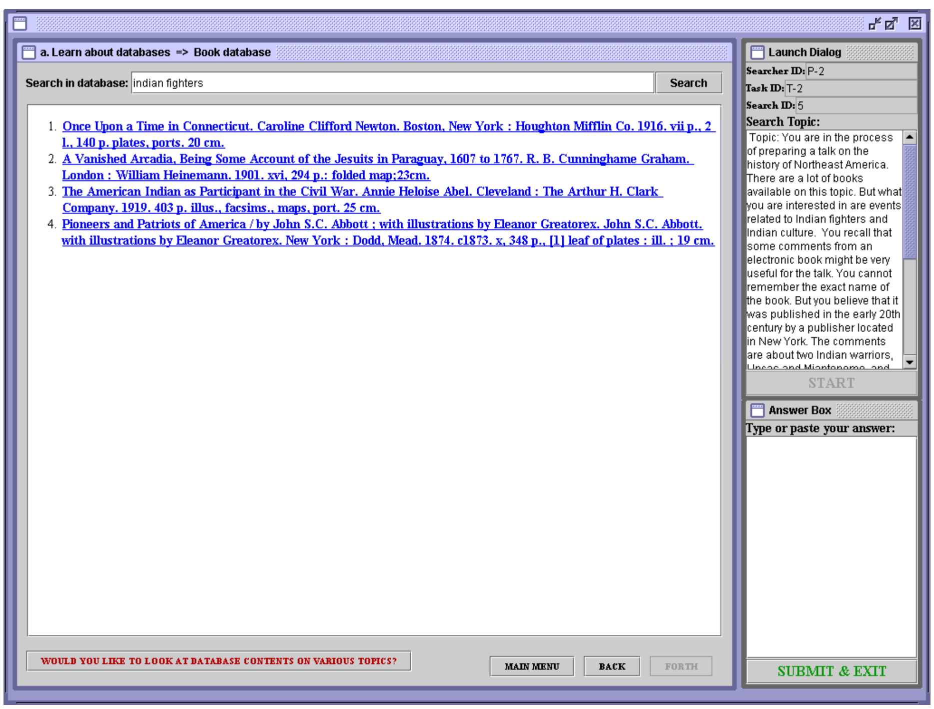


Figure 8.9 Book database contents

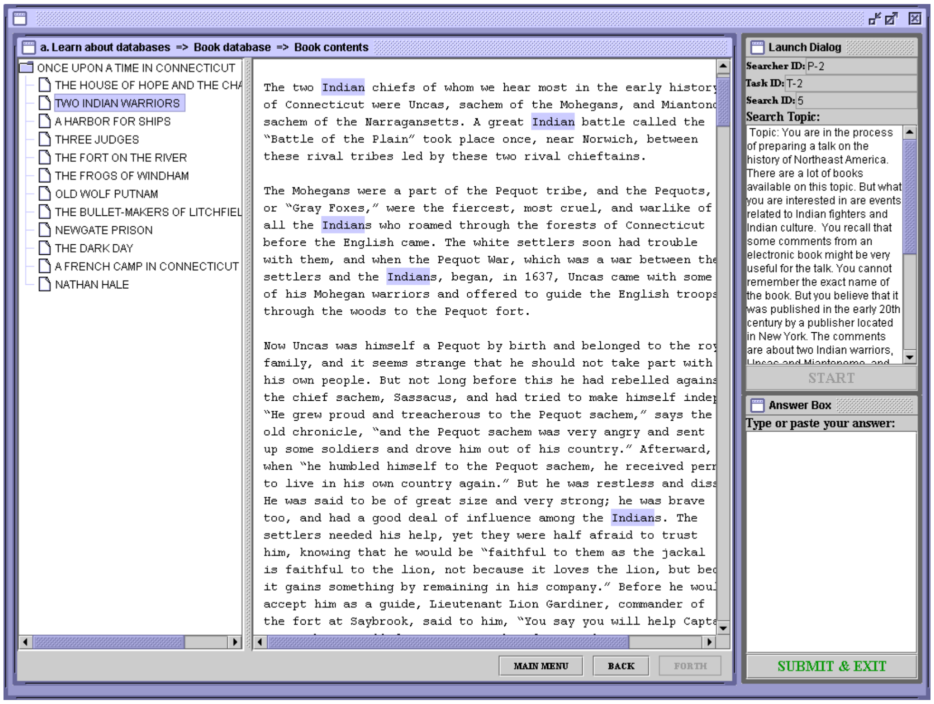


Figure 8.10 Table of contents of the book

8.4.3.3 Learn about Content Coverage of Databases on Various Topics

When the user chooses option "b. learn about content coverage of databases on various topics" from the welcome screen, s/he is directed to an interface with ten queries related to the ten experimental topics.

If the task is about finding documents, the user first selects one starting point from Figure 8.11, then clicks on the "view" button. The system shows an interface (see Figure 8.12) with a query box and a search button at the top, and the description of each database related to the query at the bottom. By using the default query or formulating new queries, the user can search a topic on one or multiple databases. The search results interfaces are similar to Figure 8.5, 8.6 and 8.7 in a sequential order.

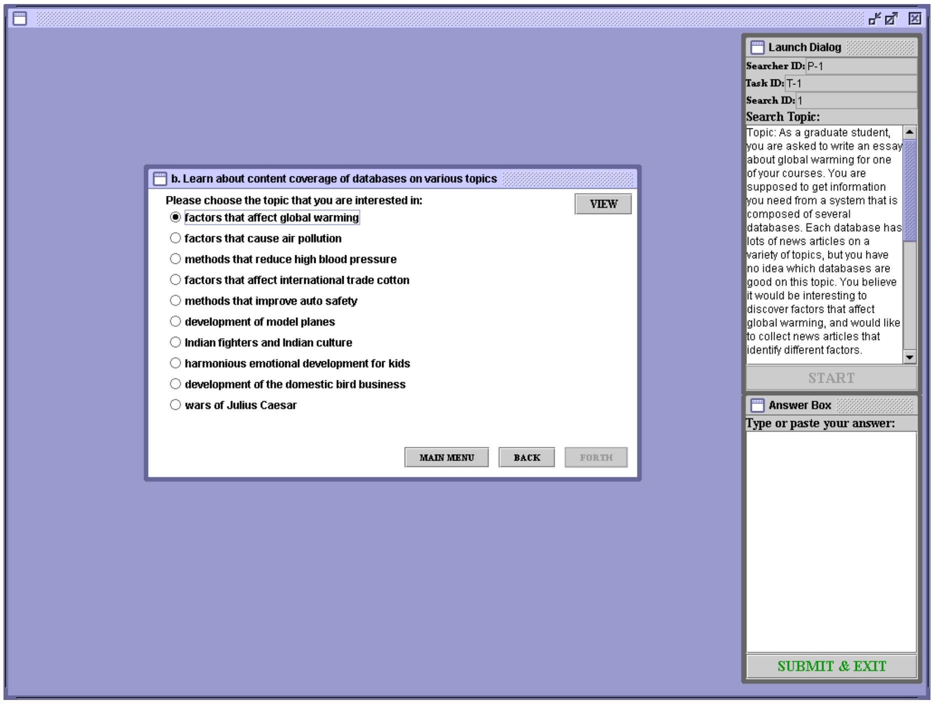


Figure 8.11 Starting points (I)

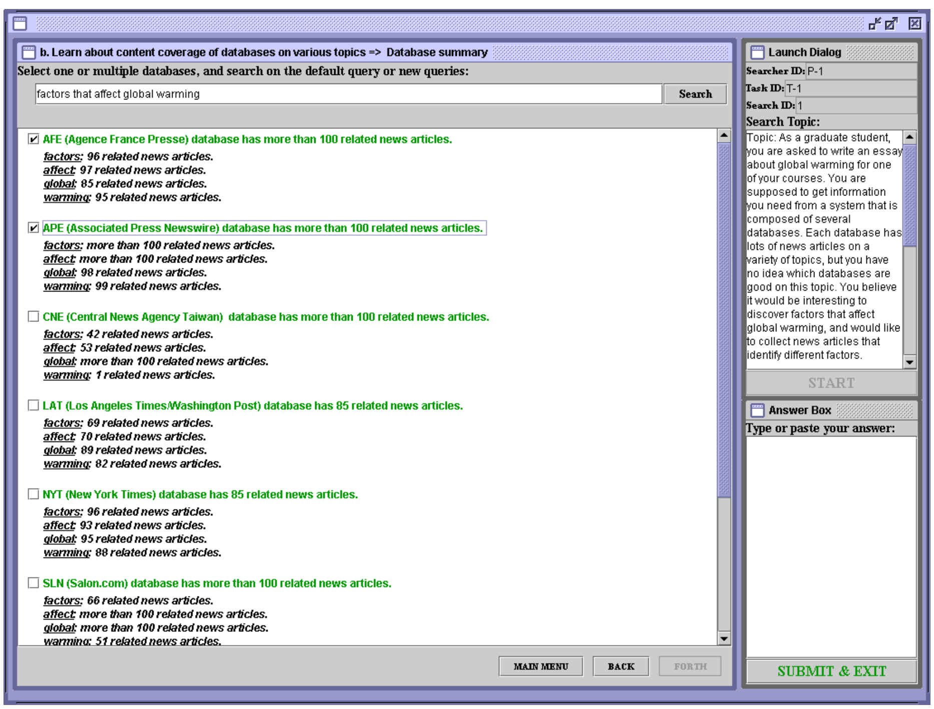
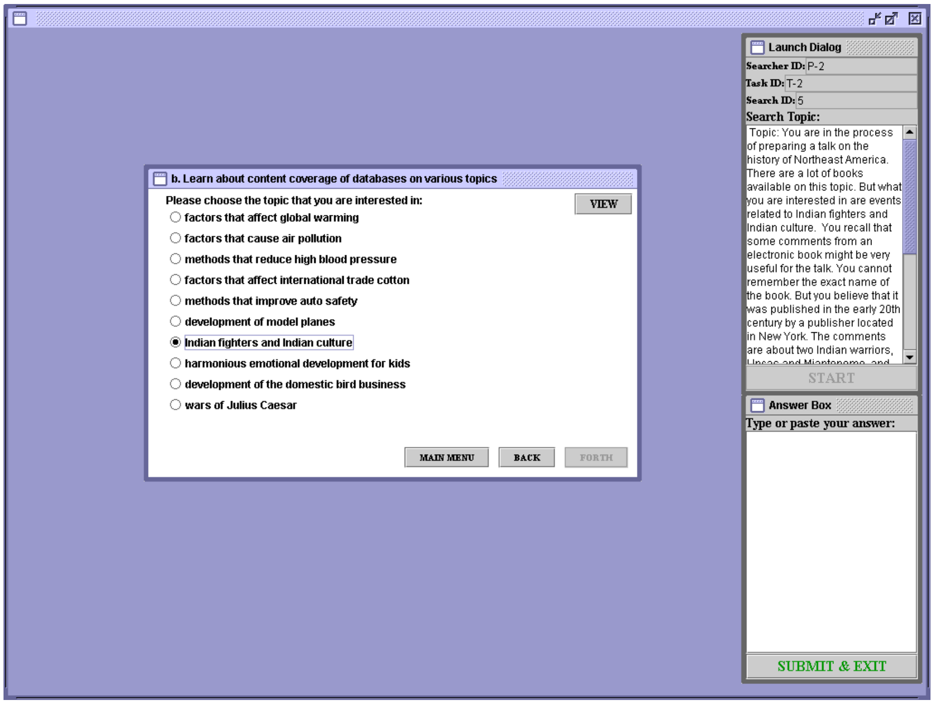


Figure 8.12 Database summary (I)

If the task is about finding books, the user first selects one starting point from Figure 8.13, then clicks on the “view” button. The system then shows an interface (see

Figure 8.14) with a query box and a search button at the top, and the description of the book database related to the query at the bottom. By using the default query or formulating new queries, the user can search a topic in this database. The search results interfaces are similar to Figure 8.9 and 8.10 in a sequential order.



<u>Figure 8.13</u> Starting points (II)

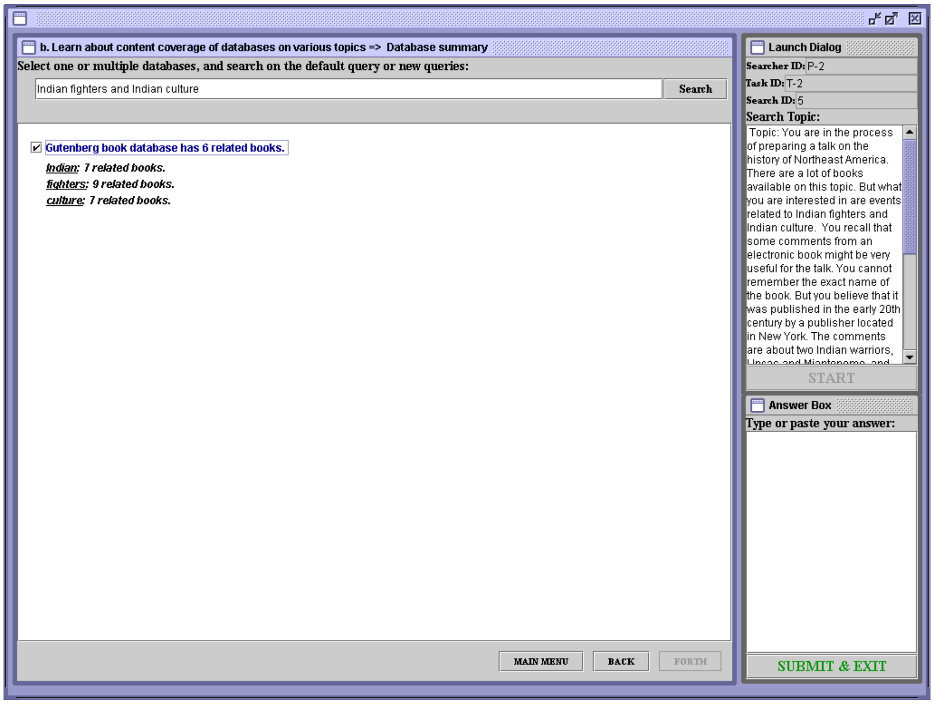


Figure 8.14 Database summary (II)

8.4.3.4 Search for Books on a Specific Topic

When the user chooses item "c. search for books on a specific topic", the screen shows as Figure 8.15. It provides several fields, such as topic, title, author, publication year, publisher and place. The user can input any query in the given fields and then start the search. The results are displayed as complete citations of books and are categorized based on the Boolean combination of different fields. For example, if the user types in a query in the place field and another query in the publication year field, then the results will be displayed as place AND publication year, place NOT publication year, publication year NOT place.

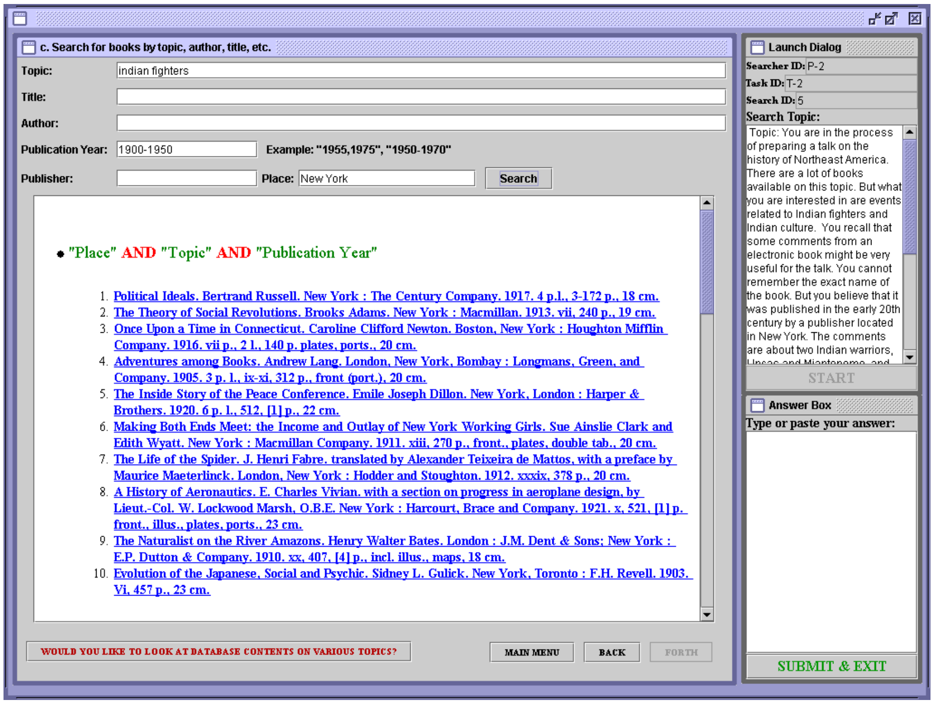


Figure 8.15 Fielded query

When the user clicks on a book, the table of contents navigation interface (similar to Figure 8.10) is displayed with the table of contents of the book at the left. As soon as the user chooses one chapter or section, the related contents are shown to the right.

8.4.3.5 Search for News Articles on a Specific Topic

When the user clicks on item "d. search for news articles on a specific topic", an interface similar to Figure 8.5 is displayed. Once the user inputs the query and clicks on the search button, the clustered retrieval results are displayed.

8.4.3.6 Other Features

The system offers the following additional features.

1. Suggestion of ISS change

This system provides a functionality that allows the user to switch between two ISSs.

Referring to Figure 8.5, at the bottom left corner of the screen there is a button labeled “would you like to look at database contents on various topics?” Clicking on this button leads the user to an interface which is similar to Figure 8.11 or 8.13, but without any marked choices. This applies to the case when the user cannot find results which satisfy her/his needs, then s/he might want to follow the suggestion to other options.

2. Spelling check

When the task is about finding books, the system provides a spelling check functionality for the author field. The input author names will be compared with the dictionary of all the valid author names in the database, and the closest match will be returned if no exact match is found. For example, if the user types in “Vivan” the system pops up a question “Did you mean Vivian for Vivan?” The user can use this function to quickly correct spelling errors to find the correct author and book s/he is looking for (see Figure 8.16). This feature is available in both baseline and experimental systems.

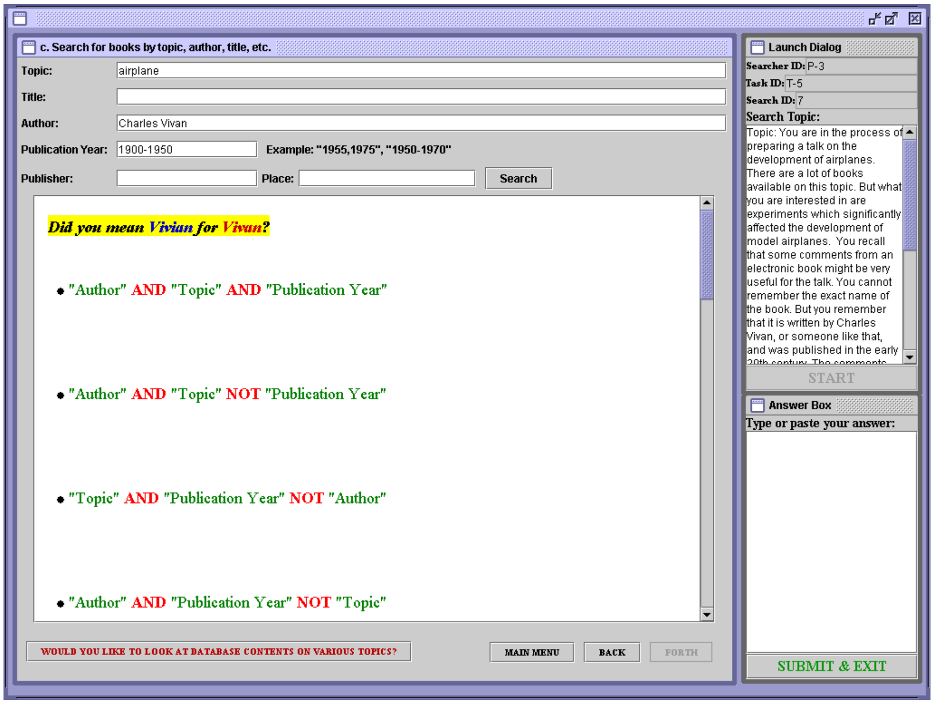


Figure 8.16 Example of spelling check

8.4.4 Baseline System Design and Implementation

The Baseline system is designed by tailoring it to the two collections (TREC HARD 2004 and Project Gutenberg). Users can either search for documents or search for comments from books in the system. It was also implemented using Java based on the LEMUR Toolkit. The system incorporates the IR support techniques used in the baseline systems in research problem 1. These IR support techniques include ranked list of documents, ranked list of complete citations of books and ranked list of paragraphs. Figure 8.17 shows the flow chart.

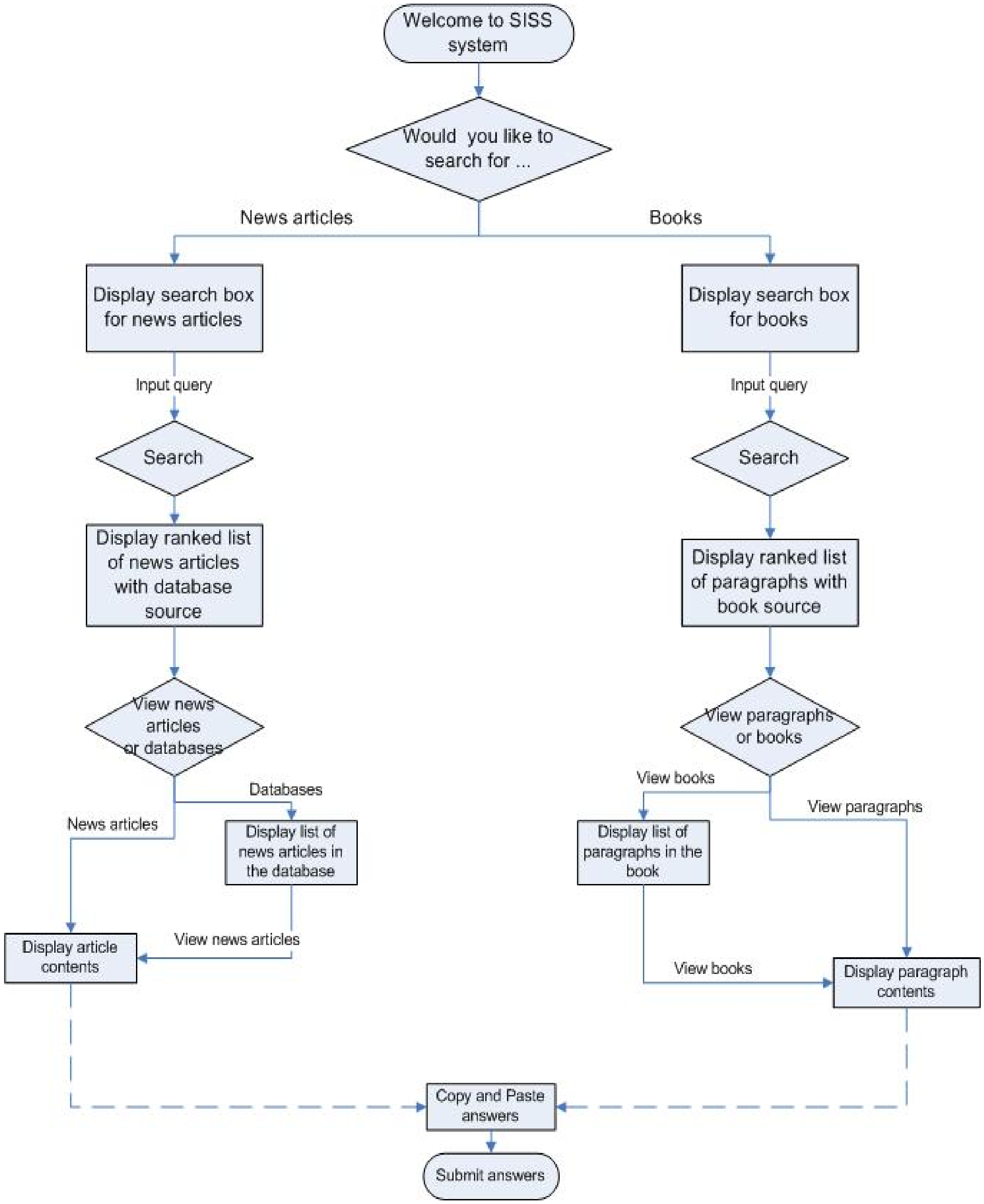


Figure 8.17 Flow chart of the baseline system

8.4.4.1 Welcome Screen

The system starts with a welcome screen: "Welcome to SISS system" (SISS stands for "Single Information-Seeking Strategies) and provides two options, searching for books on a specific topic, or searching for news articles on a specific topic (see Figure 8.18). By choosing one of these options, the user can search for information on a variety of topics. (As for MISS, eight databases are of news articles and one is a book database.)

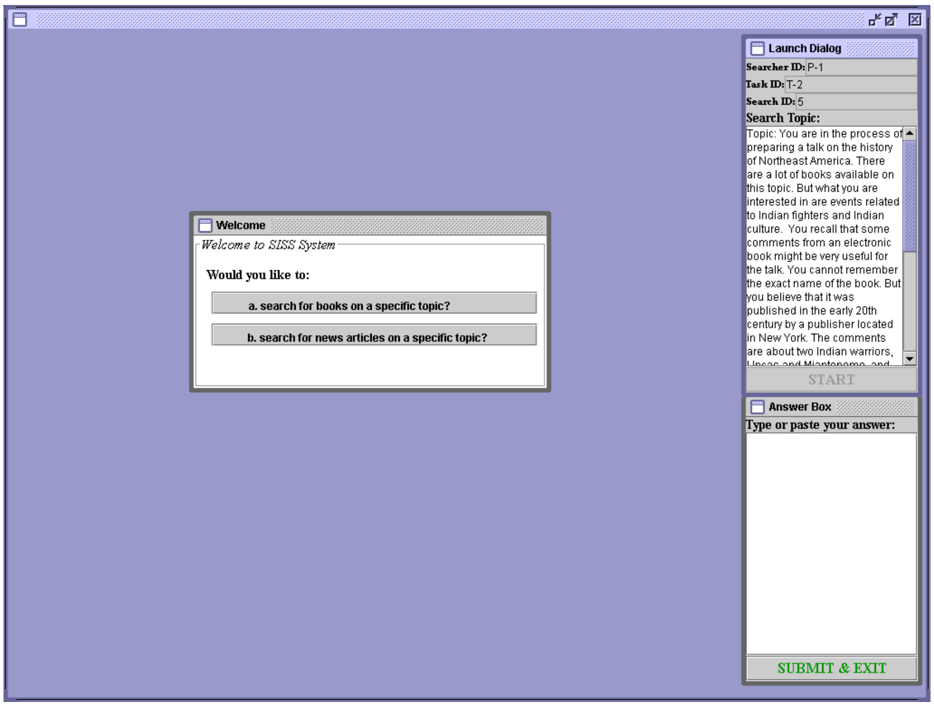


Figure 8.18 Welcome screen of the baseline system

8.4.4.2 Search for Books on a Specific Topic

At the top of the interface display, there are a query box and a search button. Once the user inputs a query and clicks on the search button, a ranked list of paragraph summaries is displayed. It also tells the source of each paragraph, which is the book that the paragraph comes from (see Figure 8.19). If the user chooses the title of the book, s/he sees the related paragraphs in this book (see Figure 8.20). If the user clicks the paragraph summary, s/he is directed to the full text of the paragraph (see Figure 8.21).

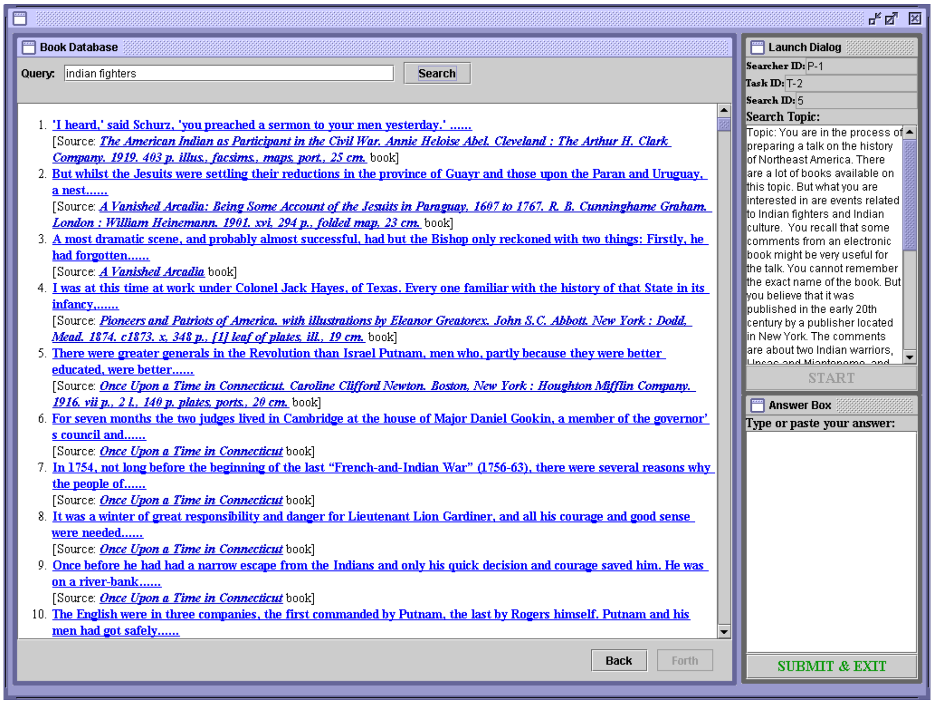


Figure 8.19 Ranked paragraphs

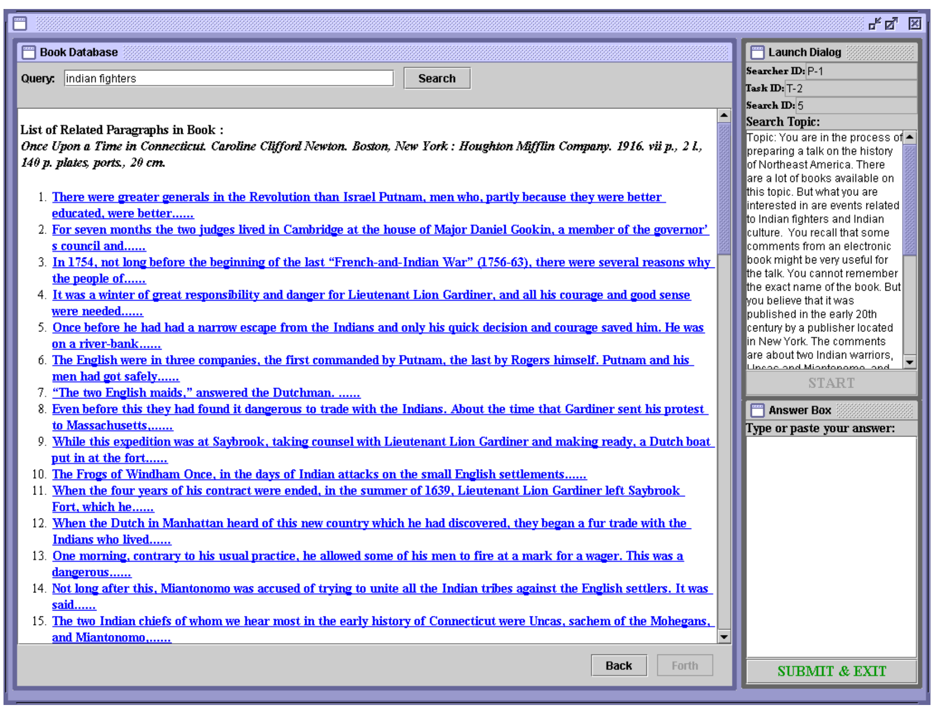


Figure 8.20 Related paragraphs in the book

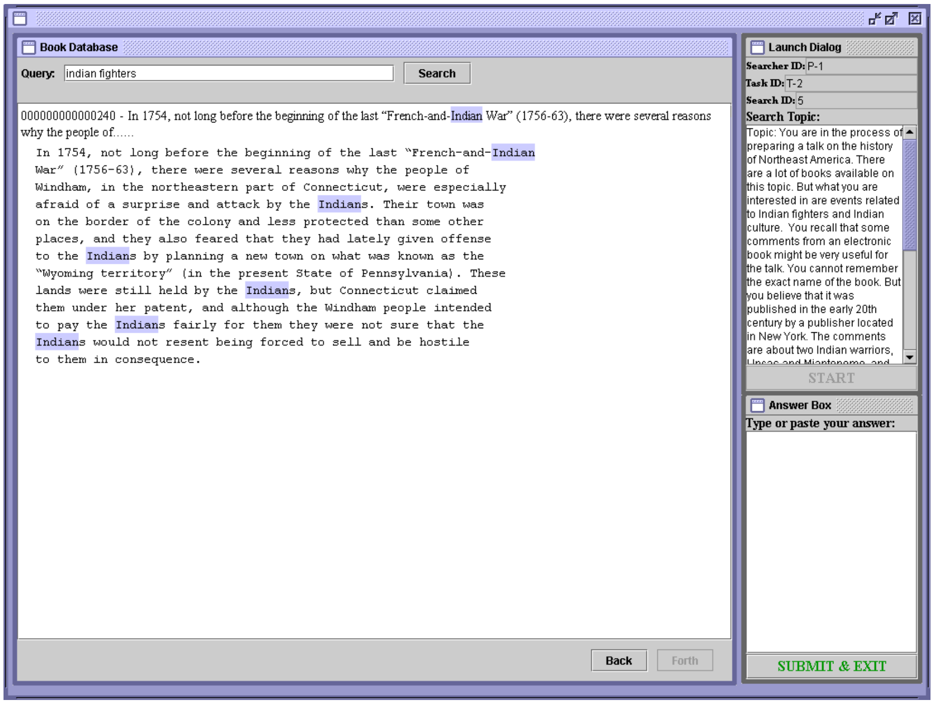


Figure 8.21 Contents of the paragraph

8.4.4.3 Search for News Articles on a Specific Topic

There are a query box and a search button at the top of the interface display. After the user inputs a query and clicks on the search button, a ranked list of documents with the source of each document is displayed (see Figure 8.22). When the user clicks on the link of each document, the detailed contents are displayed (similar to Figure 8.7). If the user chooses the database, a ranked list of documents related to this topic in the selected database is shown (see Figure 8.23).

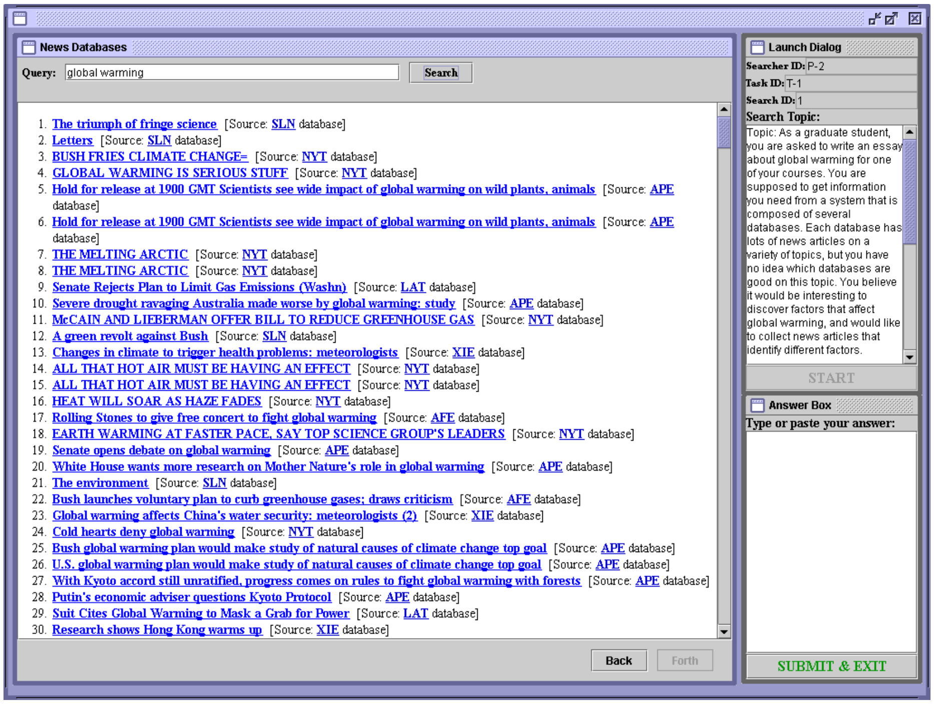


Figure 8.22 Ranked documents

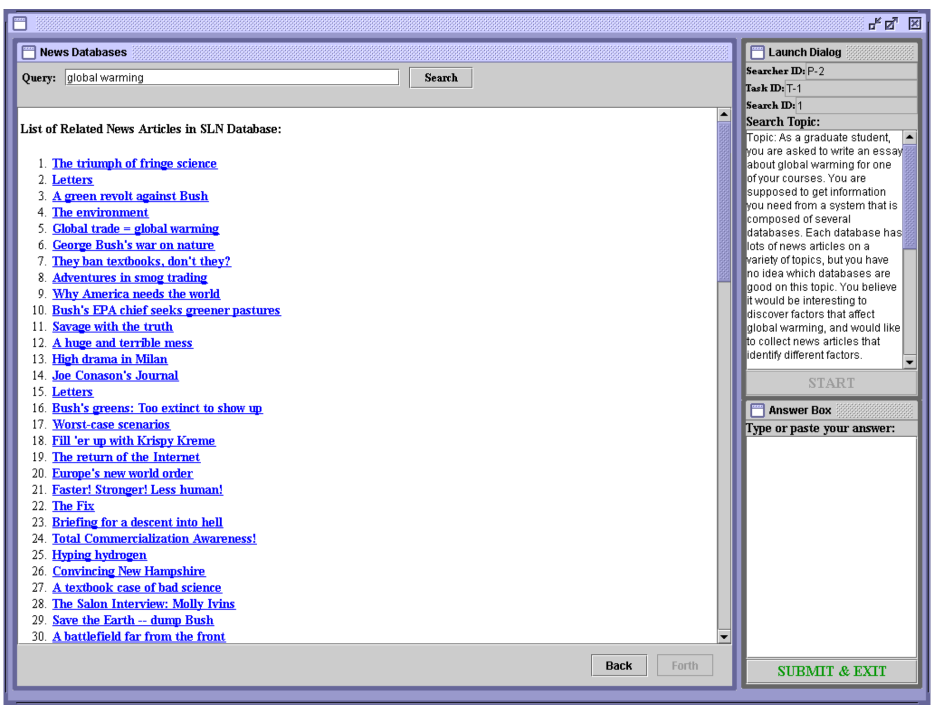


Figure 8.23 Related documents in the database

Chapter 9

Research Problem 3: Research Method

This chapter starts with the hypothesis tested in Experiment II. Then it describes the tasks and topics used in the experiment, followed by the experimental design. Finally, it discusses sampling, measures, data collection, and procedure of the experiment in detail.

9.1 Hypothesis

Hypothesis 5: An experimental system designed for supporting both scanning and searching performs better in supporting tasks requiring both scanning and searching than the baseline system designed for supporting specified searching.

9.2 Integrated Situation

Prototypical problematic situations or tasks were identified in order to lead people to engage in a variety of ISSs. Two such tasks were designed in this study. In the following, how a person might address these tasks according to our scheme of ISSs is described, and the relationship between these ISSs and corresponding IR support techniques are indicated.

These two general tasks were used as the basis for the ten specific topics (one training topic for each type of task, and four experimental topics) that were given to subjects in the experiment, and the IR support techniques were the basis of the experimental system design.

9.2.1 Scanning, then Searching

Task 1: A person is very interested in one particular topic. S/he wants to find some good documents on this topic from a system which is composed of several databases. But s/he has no idea about which of many possible databases to search.

Description: Given this situation, this person needs to first *scan* the whole system to identify the best databases for one particular topic, then conduct a *systematic search* on those databases on a specific topic. This person needs to compare the descriptions of the contents of different databases in order to choose the appropriate ones. Since s/he does not know which databases are appropriate, s/he needs to *scan* the *meta-information* of the databases in order to *recognize* the best databases. In order to get some meta-information about the databases, s/he issues a query. That query would be compared using a best match technique against the index terms associated with each database. The meta-information then is displayed by representing the database by the posting frequency of the index terms. That is, each database is summarized by the number of documents indexed by the terms, and by descriptions of the contents based on most frequent indexed terms. By doing this, the person could see how topics are covered in the databases and how they are related to each other. This representation will allow the person to compare the different databases and decide which ones look more interesting by scrolling through them.

Next, the person needs to conduct a *systematic search* within one or multiple databases for the *specific* topic in order to find documents of interest. The person needs to formulate a query based on the given task. The query is compared using a *best match technique* against the index terms associated with chosen databases. The results of the query can be represented by clustering because clustering shows the relationship among the documents, as well as the relationship between documents in the clusters and query terms or other terms that might turn out to be useful.

Query-based clusters are displayed. To accomplish this, the system needs to get the retrieval results and cluster them. Each cluster has a short summary about the documents in it and the number of documents relevant to the query topic. Now the person can decide which clusters are desirable and then scroll down to the documents within those clusters.

From the theoretical framework (see Figure 3.1), an information-seeking behavior can be seen as the movement from one ISS to another ISS. Different combinations of IR support techniques could optimally support a given type of ISS. In this task, a combination of such IR support techniques as best match, database summary, indexing and scrolling are used to support a scanning-based ISS. A combination of such IR support techniques as best match, clustering, clustered retrieval results display, and following links are used to support a searching-based ISS. The interactions between the user and the system are related to the situation, task and goal.

9.2.2 Searching, then Scanning

Task 2: A person is in the process of preparing a talk for a conference. S/he recalls some germane comments from a known electronic book but cannot remember the exact contents. S/he needs to find out the exact quotations. S/he recalls that a certain electronic book might be very helpful. But s/he cannot exactly remember the name of the book.

Description: Given this situation, the person needs to first *search* on the system to find the book, then *scan* through the book to get the comments needed. S/he has a vague recollection about a book, and needs to improve her/his knowledge of some characteristics of the book (e.g., author, title). Thus, s/he might need to search the system on terminological fragments of those data elements. In this situation, it would be good to give the person an opportunity to see something about the different characteristics about the book that s/he might remember. The items in the database, catalog or electronic books would be indexed to support a best or exact match technique within different fields such as topic, title, author, publication year, publisher, publication place. The retrieved results will be displayed as a ranked list of complete citations of books. Then the person can see the table of contents of each book by following the link of each citation.

Next, since the person has only a general idea about the quotations, s/he needs to *scan* through the *meta-information* to identify some candidate quotation page numbers. S/he might first look at *table of contents* of the book for places where the quotations might occur. Then s/he *scans* through those pages to see if the desired quotations are there and, if so, record the quotations.

In this task, a combination of such IR support techniques as table of contents navigation visualization, and scrolling are used to support a scanning-based ISS. A combination of such IR support techniques as best or exact match, indexing, fielded query search and following links are used to support a searching-based ISS. The interactions between the user and the system are related to the situation, task and goal.

9.3 Tasks and Topics

The basic idea for designing the integrated tasks is that these tasks should be an integration of the scanning and searching tasks investigated in the first experiment. For example, one task can be the integration of situation 1- task 1 (scanning – find best databases) and situation 2 - task 1 (searching - find relevant documents). Another task could be the integration of situation 1- task 2 (scanning - find book comments) and situation 2 - task 2 (searching – find the name of an electronic book). Since each task has five topics (one for training purposes and four for experimental purposes), a total of ten topics with five topics for each task were created.

<u>Integration task 1 (T1): (Situation 1 - task 1 and situation 2 - task 1)</u>

Training Topic: As a graduate student, you are asked to write an essay about air pollution for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover factors that cause air pollution, and would like to collect news articles that identify different factors.

Task: Please find as many different factors as possible. For each factor, please copy the title or link of the article which discusses that factor, and paste it to the answer box. For each article that you copy, please type or copy the factor(s) that it identifies. If an article discusses more than one factor, you only need to copy and paste the article once. If there are several articles which discuss the same factors, you only need to copy and paste one such article.

1. Topic: As a graduate student, you are asked to write an essay about global warming for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover factors that affect global warming, and would like to collect news articles that identify different factors.

Task: Please find as many different factors as possible. For each factor, please copy the title or link of the article which discusses that factor, and paste it to the answer box. For each article that you copy, please type or copy the factor(s) that it identifies. If an article discusses more than one factor, you only need to copy and paste the article once. If there are several articles which discuss the same factors, you only need to copy and paste one such article.

2. Topic: As a graduate student, you are asked to write an essay about high blood pressure for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover methods that reduce high blood pressure, and would like to collect news articles that identify different methods.

Task: Please find as many different methods as possible. For each method, please copy the title or link of the article which discusses that method, and paste it to the answer box. For each article that you copy, please type or copy the method(s) that it identifies. If an article discusses more than one method, you only need to copy and paste the article once. If there are several articles which discuss the same methods, you only need to copy and paste one such article.

3. Topic: As a graduate student, you are asked to write an essay about international trade for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover factors that affect international trade in cotton, and would like to collect news articles that identify different factors.

Task: Please find as many different factors as possible. For each factor, please copy the title or link of the article which discusses that factor, and paste it to the answer box. For each article that you copy, please type or copy the factor(s) that it identifies. If an article discusses more than one factor, you only need to copy and paste the article once. If there are several articles which discuss the same factors, you only need to copy and paste one such article.

4. Topic: As a graduate student, you are asked to write an essay about auto safety for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover methods that improve auto safety, and would like to collect news articles that identify different methods.

Task: Please find as many different methods as possible. For each method, please copy the title or link of the article which discusses that method, and paste it to the answer box. For each article that you copy, please type or copy the method(s) that it identifies. If an article discusses more than one method, you only need to copy and paste the article once. If there are several articles which discuss the same methods, you only need to copy and paste one such article.

Integration task 2 (T2): (Situation 1 - task 2 and situation 2 - task 2)

Training Topic: You are in the process of preparing a talk on the history of Northeast America. There are a lot of books available on this topic. But what you are interested in are events related to Indian fighters and Indian culture. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you believe that it was published in the early 20th century by a publisher located in New York. The comments are about two Indian warriors, Uncas and Miantonomo, and tell the story of the fate of Miantonomo. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Please find the relevant comments from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

5. Topic: You are in the process of preparing a talk on the development of airplanes. There are a lot of books available on this topic. But what you are interested in are experiments which significantly affected the development of model airplanes. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you remember that it is written by Charles Vivan, or someone like that, and was published in the early 20th century. The comments are about the first model of an airplane invented in the seventeenth century. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Please find the relevant comments from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

6. Topic: You are in the process of preparing a talk on childhood education. There are a lot of books available on this topic. But what you are interested in is the importance of harmonious emotional development for kids. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you believe that it was published in the 20th century. The comments are about why the mother's influence is important to the kid. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Please find the relevant comments from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

7. Topic: You are in the process of preparing a talk on various types of agriculture in USA. There are a lot of books available on this topic. But what you are interested in is the development of the domestic bird business. You recall that some data from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you remember that it is written by Hestings, or someone like that, and was published in the early 20th century. The data are about the relationship between poultry industry and total agricultural wealth in different states. You cannot remember the exact data, but would like to quote them in your talk.

Task: Please find the relevant data from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

8. Topic: You are in the process of preparing a talk on the history of Rome. There are a lot of books available on this topic. But what you are interested in are the wars of Julius Caesar. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you believe that it was published by a publisher in New York. The comments are about the strategies that Caesar used on the battle field to win the Battle of Pharsalia. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Please find the relevant comments from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

9.4 Experimental Design

This was a within-subjects design, in which subjects performed searches using each of the two systems, first one system, then the other. For each system, subjects first performed a search on a training topic, then searched on four different topics. So subjects searched ten topics in total. These ten topics belong to two task categories, that is, finding news-article task and finding book task, as described in 9.3. The first test topic was of the same task type as the training topic, the second topic was of the other task type, and so on. The order of the task types and topics was rotated across subjects and the experiment was replicated by exchanging the order of the two systems.

9.5 Sampling

Subjects were recruited by electronic and print postings and announcements in class. These subjects were expected to have some web search experience.

9.6 Measures and Variables

The measures and variables adopted in this experiment are the same as those in Experiment I.

9.7 Data Collection

Computer logs, questionnaires, and exit-interview were used to collect data. An entry questionnaire (Appendix C(2)) gathered demographic and other background information. A pre-search questionnaire (Appendix C(3)) elicited information about subjects' knowledge of the topic. A post-search questionnaire (Appendix C(4)) collected opinions about the particular search. A post-system questionnaire (Appendix C(5)) elicited opinions about the specific system. An exit interview (Appendix C(6)) collected opinions of the two systems and the whole search process.

Logging software, "Techsmith Morae 1.3" (http://www.techsmith.com/morae.asp) was used to log the interaction between the user and the system, as well as to record what the user said during the whole experiment. Morae software can record and synchronize data of user and system for usability analysis. It contains three parts: Recorder that can capture the interaction between the user and the system in video and data format; Remote Reviewer that can allow multiple observers to view or hear the interaction activity; Manager that can help input the recorded data and perform data analysis. This software was used because it can capture very rich meta-information, and it is easy for post data processing.

9.8 Procedure

When subjects arrived, they completed an informed consent form (Appendix C(1)), which included detailed instructions about the experiment, and then the entry questionnaire. Next, they were given a training topic to practice with the first system they would use, then they searched on four topics using the first system. For each topic, they filled out a pre-search questionnaire, conducted the search and saved the answers in the given place. When they felt that a satisfactory answer was saved, or they ran out of time (subjects had up to 12 minutes per search), they answered a brief post-search questionnaire. This procedure continued until four topics in the first system were completed, after which they filled out a post-system questionnaire and were given a three-minute break. The same procedure was followed for the next set of topics using the second system, after which the exit interview was given. Each subject was paid $30 cash equivalent value (gift card/cash) after completing the experiment. Figure 9.1 shows the procedure.

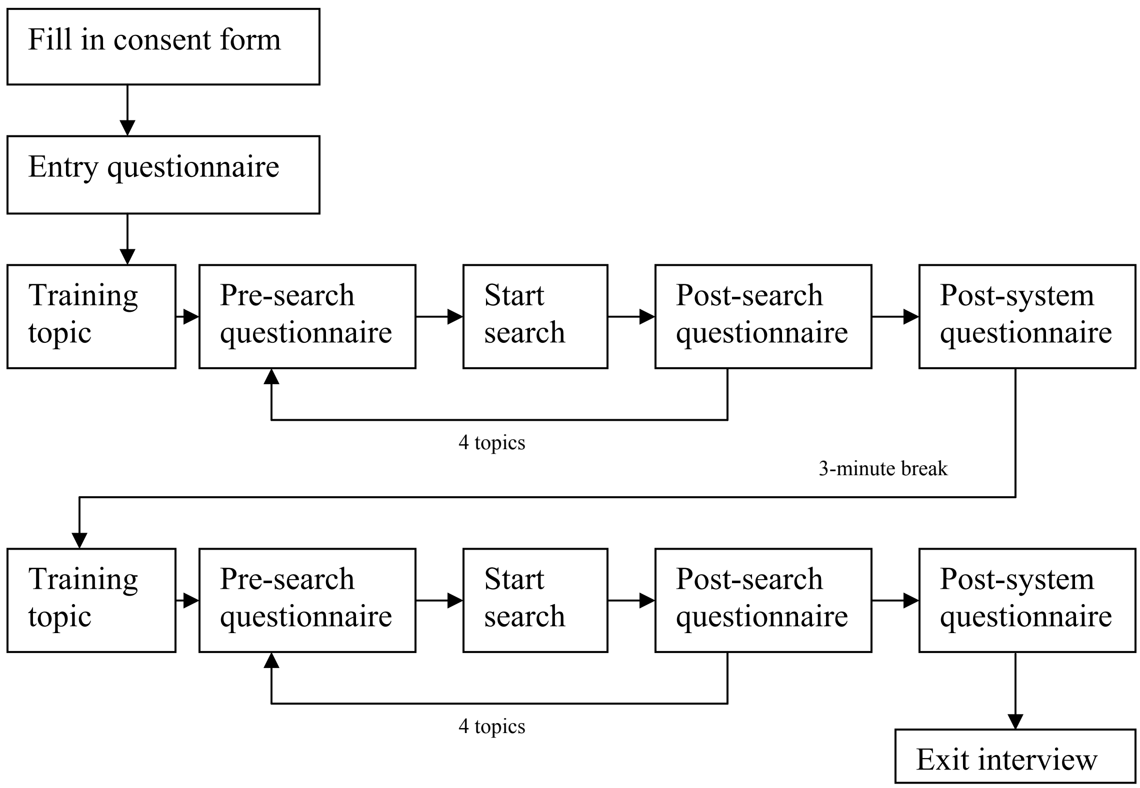


Figure 9.1 Experimental procedure (Experiment II)

Chapter 10

Research Problem 3: Results

This chapter starts with the pilot results of the experiment II and then follows with the subjects' characteristics and their computer and searching experience. A description of the performance and interaction measures of the systems follows. Next, the data from pre-search, post-search and post-system questionnaires are presented. Finally, the results from the exit interview are reported.

10.1 Pilot Results of Experiment II

Four subjects were recruited in this study. These subjects were all Rutgers graduate students.

10.1.1 Systems and Tasks

None of these subjects had any problems understanding the given tasks and the systems. The systems and tasks were proved to be valid by the users' responses to the questions and the pilot results.

10.1.2 Preliminary Findings

10.1.2.1 Usability of the Systems

Results from post-system questionnaires showed that subjects felt the experimental system was more understandable and more useful than the baseline system, but less easy to learn to use and less easy to use than the baseline system. Since all the subjects favored Google, the results seemed reasonable. The ranked list of paragraphs or documents was quite similar to Google, which made the baseline system easier for them to learn and use.

In the exit interview, subjects showed strong preferences toward the experimental system. Most of them believed the experimental system was more helpful, easier to use and more capable of conducting better searches.

10.1.2.2 Features Subjects Liked Most

In book search tasks, subjects using the experimental system liked the fielded query feature. Subjects using the baseline system mentioned that it was necessary to have a place for them to search for a book by author, publication year, and other fields.

In document search tasks, subjects using the experimental system liked the feature of database summary. They found it useful to know about the structure of the databases from the description of how many documents were related to the terms in a query. They also liked the ranked cluster representation.

Subjects thought it was a good idea to allow them to search on one or multiple databases. In that way, they had better control on searched database sources.

10.1.2.3 Features Subjects Disliked Most

Most subjects didn’t like ranked paragraphs or ranked documents in the baseline system because it was hard for them to narrow down their searches to a small number and to make wise choices on the most relevant documents or paragraphs.

In summary, the pilot study demonstrated that the experimental design was successful and effective. In addition, the results showed the value of the experimental system.

10.2 Results of Experiment II

10.2.1 Subjects

Thirty-four subjects participated in this study. Data from two subjects had to be discarded because the subjects had trouble fully understanding the instructions and tasks. So effectively there were 32 subjects.

Among these subjects, 24 (75%) were female and 8 (25%) were male, with age from 22 to 59. Most of the subjects (68.8%) were younger than 30. 44% of the subjects were master's students, and 56% were Ph.D. students who had earned a master's degree already. On average, the subjects had 7.34 years experience of web searching.

Subjects' computer and searching experience were collected from the entry questionnaire (Appendix C(2)). Subjects were asked to indicate their computer and searching experience on a 7-point scale, where 1 = "none"; 4 = "some"; and 7 = "a great deal." Subjects were asked to indicate their level of expertise with computers or searching on a 7-point scale, where 1 = "novice" and 7 = "expert." Subjects were also asked about the frequency of their searching on a 7-point scale, where 1 = "never"; 4 = "monthly"; and 7 = "daily." Besides, subjects were asked to indicate whether they can usually find what they were looking for on a 7-point scale, where 1 = "rarely"; 4 = "sometimes"; and 7 = "often."

These results are listed below in Table 10.2. Subjects had very frequent use of computers (M=6.91, SD=0.39), high expertise of computer (M=5.34, SD=1.15), high searching experience of catalog (M=5.63, SD=1.29), and WWW (M=6.72, SD=0.58), very high frequency of search (M=6.50, SD=0.92), high expertise of searching (M=5.28, SD=0.58), and very high confidence in finding what they need from searching (M=6.06, SD=0.95). Subjects also had long-term period of searching (M=7.34 years, SD=2.24 years). However, their searching experience of commercial systems was relatively low (M=3.88, SD=1.91).

Table 10.1

Subject Characteristics (Experiment II)

Characteristics	Values	No. of subjects
Age	<30	22
	30-39	3
	40-49	4
	>=50	3
Current major	Library and information studies	9
	Communication	5
	Computer science	3
	Political science	3
	Anthropology	2
	Biomedical engineering	2
	Others	8
Degree earned	Ph.D.	
	Master	18
	Bachelor	14

Table 10.2

Computer and Searching Experience of Subjects (Experiment II)

Demographic data	Mean (standard deviation)
Computer daily use	6.91 (0.39)
Expertise of computer	5.34 (1.15)
Searching experience of catalog	5.63 (1.29)
Searching experience of commercial system	3.88 (1.91)
Searching experience of WWW	6.72 (0.58)
Searching experience of other systems	1.14 (0.38)
Frequency of search	6.50 (0.92)
Search information found	6.06 (0.95)
Expertise of searching	5.28 (0.58)
Number of years of searching experience	7.34 (2.24)

10.2.2 Performance

Time of task completion, user satisfaction with the results of their search, result correctness, and aspectual recall were the performance measures. Time was collected by system logs, with the start point being the time when the user pressed the "START" button and started to search, and the end point being the time when the user submitted the answers and exited the system. User satisfaction was assessed by post-search questionnaire (Appendix C(4)). Result correctness, the performance measure for the book tasks, was measured as the experimenter's rating of whether a search resulted in a saved paragraph which answered the search topic as measured on a 3-point scale: *Incorrect(wrong book), Partially Correct(wrong paragraph, right book),* and *Correct(right paragraph)*. Subjects were asked to rate their satisfaction with their search results on a 7-point scale, where 1 = "not at all"; 4 =”somewhat"; 7 = "extremely." Aspectual recall, the performance measure for the article tasks, was determined by pooling all of the aspects identified for each topic by all of the subjects. Aspectual recall, a measure developed in the TREC Interactive Track (Dumais & Nick, 2005), is the ratio of aspects of the search topic identified in the documents saved by the subject, to the total number of aspects of the topic identified by all the subjects in the experiment. In this specific study, the experimenter was the assessor. Different subjects may use different wordings for the similar aspects. The experimenter interpreted the aspects identified by the subjects, and grouped them into broader categories. For example, for the factors that could improve automobile safety, several subjects identified "side airbags", "head airbags", "front airbags" and "smart airbags" as the factors, which were all grouped into the aspect of "airbags."

Table 10.3

Performance of Systems (Experiment II)

Systems	Mean (standard deviation)			
	Time (mins)	Result satisfaction (1-7)	Result correctness (0-2)	Aspectual recall
Baseline	8.94 (3.05)	4.86 (1.77)	0.97 (0.84)	0.44 (0.21)
Experimental	9.11 (2.91)	5.40 (1.43)	1.17 (0.77)	0.54 (0.21)

Table 10.4

Significance Value of Systems (Experiment II)

Systems	ANOVA		Chi-square	Wilcoxon signed rank
	Time	Aspectual recall	Result correctness	Result satisfaction
Baseline / Experimental	0.657	0.009*	0.213	0.008*

*$p < 0.01$

In Table 10.3 & Table 10.4, results from ANOVA indicated that subjects using the experimental system spent only slightly more time (M =9.11, SD=2.91) than subjects using the baseline system (M =8.94, SD=3.05), but the difference was not significant, $F(1,254)=0.198$, $p=0.657$. Pearson chi-square test showed that there was no significant relationship between system and result correctness, although subjects using the experimental system found somewhat more correct answers (M =1.17, SD=0.77) than subjects using the baseline system (M =0.97, SD=0.84), $\chi^2 =3.093$, df=2, $p=0.213$ (the number of incorrect, partially correct and correct answers by system is shown in Table 10.6). Wilcoxon signed rank test results showed that subjects felt significantly more satisfied with their results using the experimental system (M =5.40, SD =1.43) than the baseline system (M =4.86, SD =1.77), $Z=-2.633$, $p=0.008$. ANOVA results showed that subjects using the experimental system identified significantly more relevant aspects (M =0.54, SD=0.21) than those using the baseline system (M =0.44, SD=0.21), $F(1, 126)=6.951$, $p=0.009$.

By looking at the time and result satisfaction by task type, as shown in Table 10.5, results indicated that the subjects spent much less time on the book tasks (M =7.29, SD =3.11 for the baseline system, and M =7.35, SD =2.59 for the experimental system) than the article tasks (M =10.59, SD =1.87 for the baseline system, and M =10.86, SD =2.03 for the experimental system), and the result satisfaction was slightly less for the book tasks (M =4.61, SD =1.97 for the baseline system, and M =5.36, SD =1.71 for the experimental system) than the article tasks (M =5.11, SD =1.51 for the baseline system, and M =5.44, SD =1.08 for the experimental system). The overall observation was still true for both book and article tasks, i.e., the subjects spent slightly more time on the experimental system than the baseline system, and were more satisfied with the results of the experimental system than the baseline system.

Table 10.5

Time and Result Satisfaction by Task Type (Experiment II)

Tasks	Systems	Time (mins)	Result satisfaction (1-7)
Book	Baseline	7.29 (3.11)	4.61 (1.97)
	Experimental	7.35 (2.59)	5.36 (1.71)
Article	Baseline	10.59 (1.87)	5.11 (1.51)
	Experimental	10.86 (2.03)	5.44 (1.08)

Since time was also a very important measure for efficiency, a boxplot (see Figure 10.1) is used to show the distribution of time across both systems by task type. Subjects using the experimental system spent more time than those using the baseline system, but not significantly so.

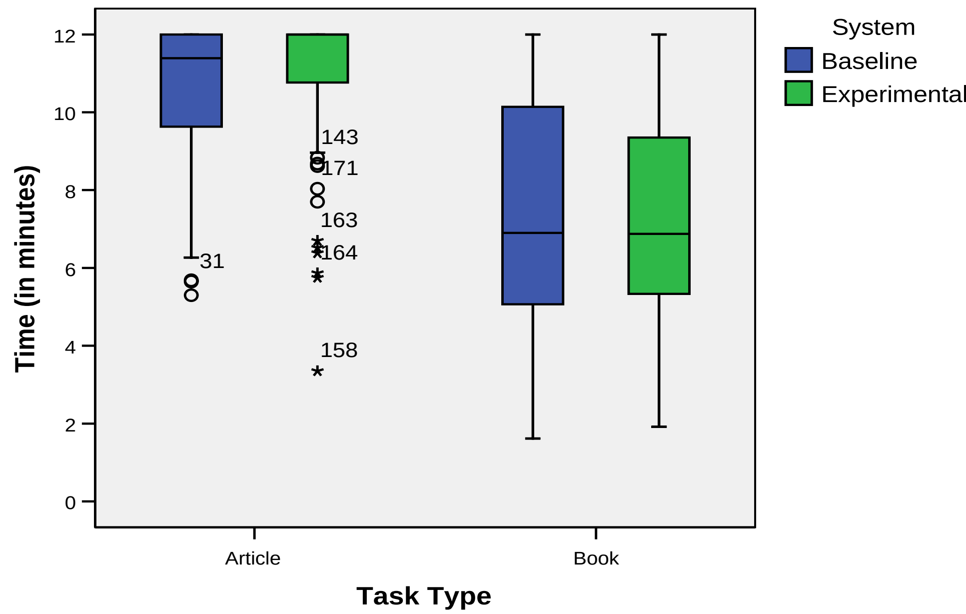


Figure 10.1 Time distributions across systems (Experiment II)

Table 10.6

Result Correctness across Systems (Experiment II)

Systems	Result correctness		
	Incorrect (0)	Partially correct (1)	Correct (2)
Baseline	23	20	21
Experimental	14	25	25

10.2.3 Interaction

Table 10.7 shows definitions of the interaction variables for the respective systems, including number of iterations, number of final saved documents/paragraphs, number of documents/books viewed and query length (see Table 10.8 for the results for each of these variables).

Table 10.7

Variables Used to Describe Search Behavior of Interaction (Experiment II)

Variables	Definitions
Number of iterations	The total number of searches during the entire search process (a search was identified by a query submitted by the user)
Number of final saved documents/paragraphs	The total number of documents/paragraphs which were saved by the searcher at the end of the search
Number of documents/books/paragraphs viewed	The total number of documents/books/paragraphs whose contents were displayed to the searcher
Query length	Length of query (total number of words in a query, or total number of words in all the fields for fielded query)

Table 10.8

Mean and Standard Deviation of Interaction Variables (Experiment II)

Interaction Variables	Systems	
	Baseline	Experimental
Number of iterations	3.81 (3.65)	2.96 (2.68)
Number of final saved documents/paragraphs	4.55 (2.22)	4.58 (2.37)
Number of documents/books/paragraphs viewed	7.98 (4.71)	9.64 (8.75)
Query length	3.39 (1.20)	4.78 (2.39)

From Table 10.8, ANOVA results showed that subjects using the experimental system had significantly less iterations (M=2.96, SD =2.68) than those using the baseline system (M =3.81, SD =3.65), $F(1,254)=4.516$, $p=0.035$. Subjects in both systems saved almost exactly the same number of documents/paragraphs, on average. ANOVA results showed that subjects using the experimental system viewed more documents/books/paragraphs (M =9.64, SD =8.75) than those using the baseline system

(M =7.98, SD =4.71), but the difference was not significant, $F(1,254)=3.588$, $p=0.059$. Subjects using the experimental system employed significantly longer queries (M =4.78, SD =2.39) than the baseline system (M =3.39, SD =1.20), $F(1,254)=34.571$, $p<0.001$.

ANOVA tests were performed to test whether fielded queries were the cause of the significantly longer queries in the experimental system. Results indicated that fielded queries were significantly longer (M=5.68, SD=2.52) than non-fielded queries in the experimental system (M=4.01, SD=0.98), $F(1,62)=12.03$, $p=0.001$. Fielded queries in the experimental system were significantly longer (M=5.68, SD=2.52) than queries in the baseline system (M=3.39, SD=1.20), $F(1,62)=24.37$, $p<0.001$. Non-fielded queries in the experimental system were significantly longer (M=4.01, SD=0.98) than queries in the baseline system (M=3.39, SD=1.20), $F(1,62)=8.53$, $p=0.005$.

MANOVA tests were conducted to test how the number of unique words used in the queries differs for each topic. The results indicated that significantly more unique words were used in topic 6 (childhood education) than any other topic. None of the other topics had significantly different unique words from each other. A boxplot chart (see Figure 10.2) of the number of unique words for all the topics is as below. For topic 6, there are two data points that fall outside of the chart, which have 59 and 76 unique words respectively. It turns out that two subjects pasted whole sentences into their queries.

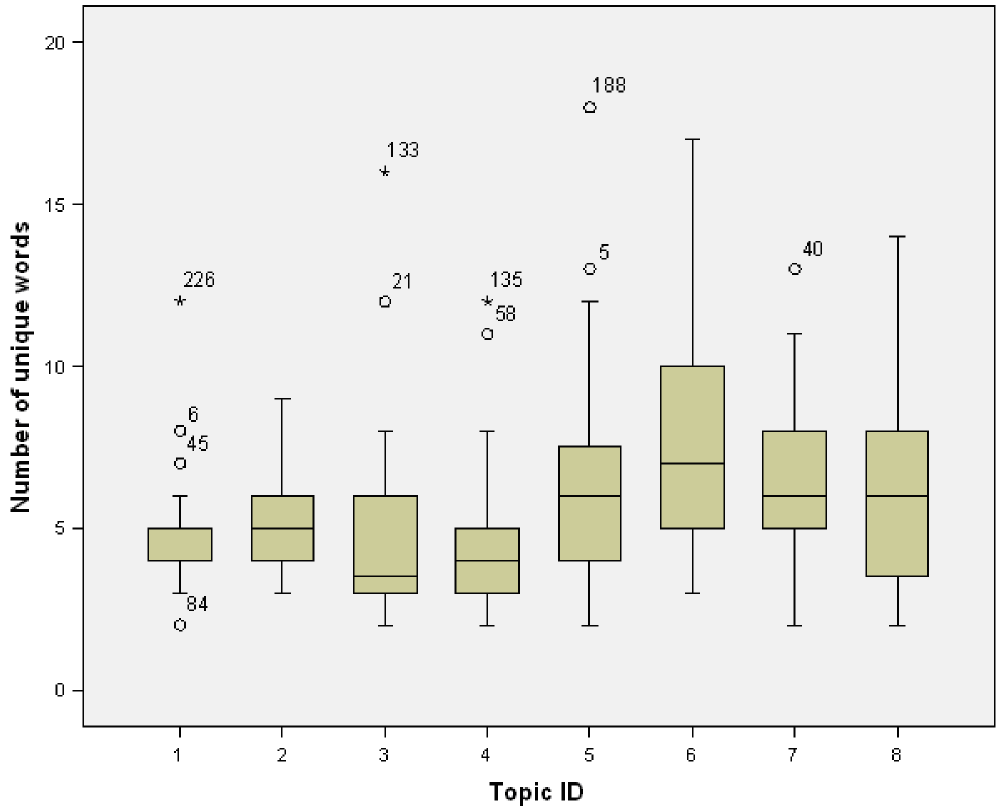


Figure 10.2 Number of unique words across topics (Experiment II)

10.2.4 System Order and Task Order Effect

A one-way ANOVA test was employed to test whether the order of system usage between experimental and baseline condition had an impact on time. The results showed that those in EB system order group (experimental first) spent somewhat more time (M=9.15, SD=2.95) than those in BE system order group (the baseline system first) (M=8.89, SD=3.00), but not significantly so, $F(1,254)=0.502$, $p=0.479$. The test of homogeneity of variance (Levene's test) was not significant, $p=0.962$, which showed that the variances of two groups were not significantly different.

An ANOVA test was employed to test whether the order of system usage between experimental and baseline condition had an impact on result correctness. The results showed that those in EB system order group had slightly higher result correctness (M=1.19, SD=0.81) than those in BE system order group (M=0.95, SD=0.79), but not significantly so, $F(1,126)$=2.748, p=0.100. The test of homogeneity of variance (Levene's test) was not significant, p=0.262, which showed that the variances of two groups were not significantly different.

A 1 x 4 ANOVA test was employed to test whether the order of task had an impact on time. The results showed that task 1 (M=9.02, SD=2.90) and task 2 (M=9.60, SD=2.75) used more time than task 3 (M=8.63, SD=3.30) and task 4 (M=8.84, SD=2.90), but not significantly so, $F(3,252)$=1.239, p=0.296. The test of homogeneity of variance (Levene's test) was not significant, p=0.194, which showed that the variances of four groups were not significantly different. Post-hoc comparisons were also performed using Scheffe test, with no significant results between any two groups.

A 1 x 4 ANOVA test was employed to test whether the order of task had an impact on result correctness. The results indicated that task 1 (M=1.13, SD=0.75), task 2 (M=1.13, SD=0.75), and task 3 (M=1.13, SD=0.83) had higher correctness than task 4 (M=0.91, SD=0.89), but not significantly so, $F(3,124)$=0.584, p=0.626. The test of homogeneity of variance (Levene's test) was not significant, p=0.198, which showed that the variances of four groups were not significantly different. Post-hoc comparisons were also performed using Scheffe test, with no significant results between any two groups.

MANOVA (2x4) was conducted to evaluate whether the interaction between system order and task order had any effect on time, results correctness and aspectual recall. The results were not significant, with $F(3,248)$=0.320 and p=0.811 for effect on time, F(3,120)=0.522 and p = 0.668 for effect on result correctness, F(3,120)=0.514 and p=0.673 for effect on aspectual recall.

10.2.5 Pre-search Questionnaire

In the pre-search questionnaire (Appendix C(3)), subjects were asked about their familiarity with the given topic on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely." Subjects were asked to indicate their level of expertise with the given topic on a 7-point scale, where 1 = "novice"; and 7 = "extremely."

Table 10.9 shows the mean and standard deviation of these two variables across the topics.

Table 10.9

Topic Familiarity and Expertise (Experiment II)

Task No.	Topics	Mean (standard deviation)	
		Topic familiarity	Topic expertise
Training-book	History of America	1.56 (0.91)	1.41 (0.71)
Training-article	Air pollution	3.72 (1.42)	3.00 (1.32)
1	Global warming	4.22 (1.29)	3.44 (1.39)
2	High blood pressure	3.31 (1.28)	2.66 (1.36)
3	International trade in cotton	1.91 (1.23)	1.78 (1.10)
4	Auto safety	2.94 (1.32)	2.47 (1.14)
5	Development of airplane models	1.88 (1.07)	1.56 (0.84)
6	Childhood education	2.69 (1.69)	2.16 (1.32)
7	Development of the domestic bird business	1.63 (0.94)	1.50 (0.80)
8	History of Rome	2.16 (1.37)	1.84 (1.17)

Table 10.9 shows that subjects' mean self-reported expertise and familiarity were all pretty uniformly low, for all topics, with rather low standard deviation, as well.

10.2.6 Post-search Questionnaire

Subjects' opinions about each task were assessed by a post-search questionnaire (Appendix C(4)). Subjects were asked whether it was easy to get started on the search, whether it was easy to do the search on the specific topic, whether they were satisfied with the results, and whether they had enough time to do the search on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely." Table 10.10 lists the mean and standard deviation for each system for these questions. The systems were compared for each question based on each subject's responses using the non-parametric Wilcoxon signed rank test. SPSS 14.0 software was used to do the analysis.

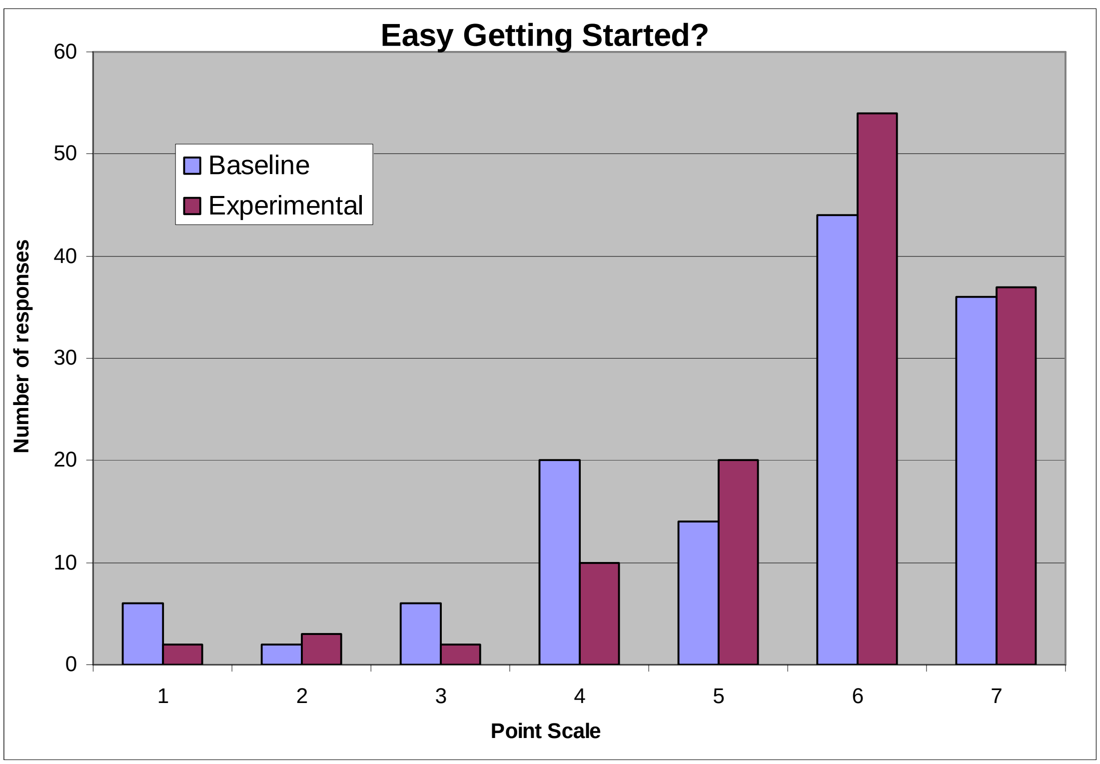
Results (see Table 10.10) showed that it was significantly easier to get started using the experimental system (M=5.76, SD=1.27) than using the baseline system (M=5.42, SD=1.60), Z=-2.239, p=0.025. Subjects felt it was easier to search using the experimental system (M=5.37, SD=1.32) than using the baseline system (M=5.05, SD=1.65), but the difference was not significant, Z=-1.341, p=0.180. Subjects were significantly more satisfied with results when using the experimental system (M=5.40, SD=1.43) than using the baseline system (M=4.86, SD=1.76), Z=-2.633, p=0.008. Subjects felt that they had significantly more enough time to do the search when using the experimental system (M=5.93, SD=1.27) than using the baseline system (M=5.51, SD=1.65), Z=-2.466, p=0.014. See Figure 10.3 for the distribution of responses to the post-search questionnaire.

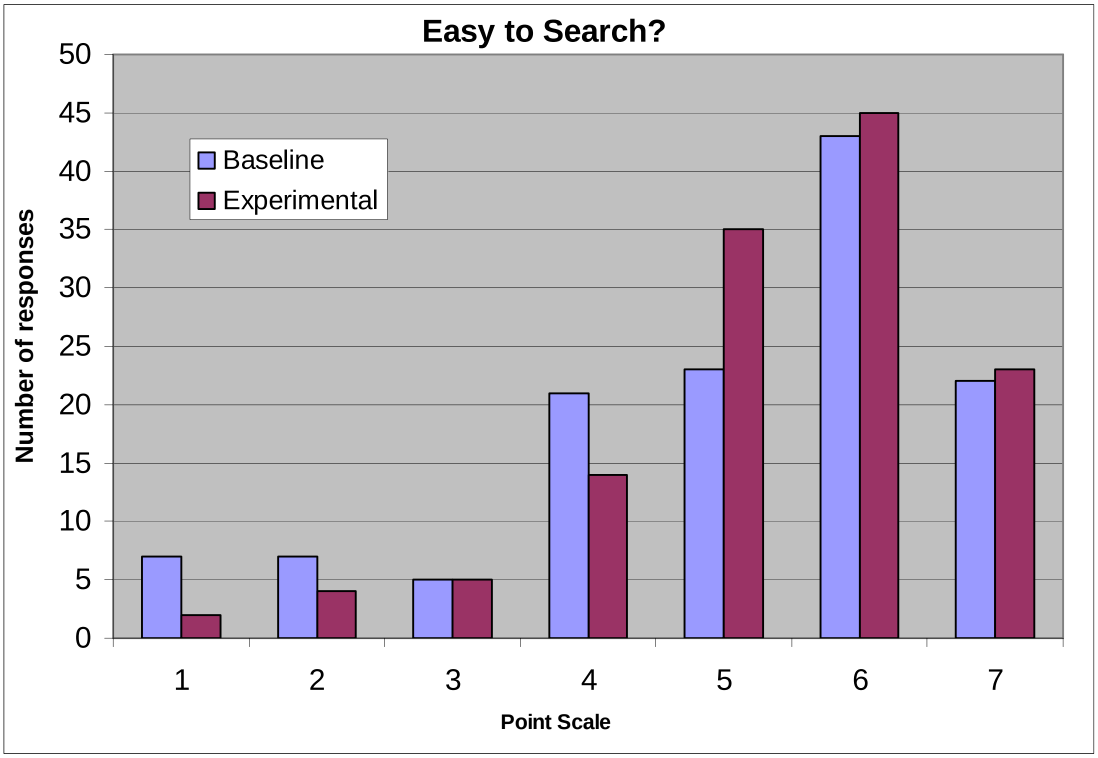
Table 10.10

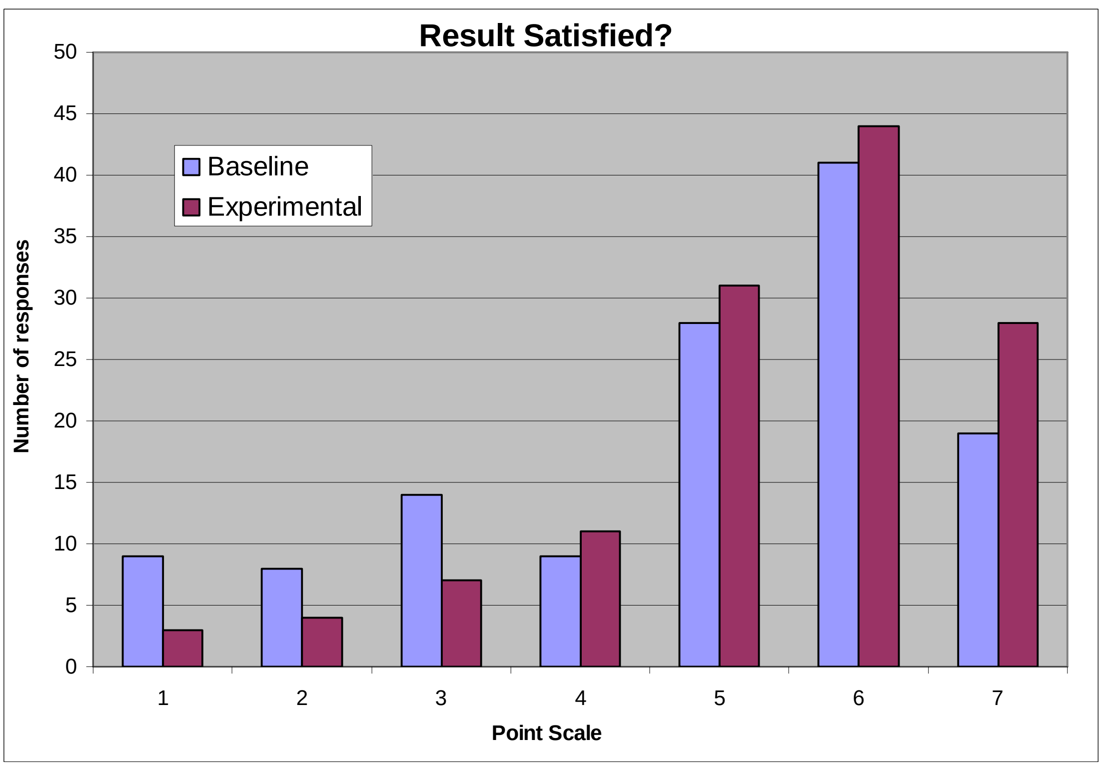
Post-search Questionnaire Results (Experiment II)

	Systems	Easy getting started?	Easy to search?	Result satisfied?	Enough time?
Mean	Baseline	5.42	5.05	4.86	5.51
	Experimental	5.76	5.37	5.40	5.93
Standard deviation	Baseline	1.60	1.65	1.76	1.65
	Experimental	1.27	1.32	1.43	1.27
p-value	Baseline Experimental	0.025*	0.180	0.008*	0.014*

*$p < 0.05$




Easy to Search?
Baseline
Experimental
Number of responses
0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7
Point Scale


Result Satisfied?
Baseline
Experimental
Number of responses
0
5
10
15
20
25
30
35
40
45
50
1
2
3
4
5
6
7
Point Scale

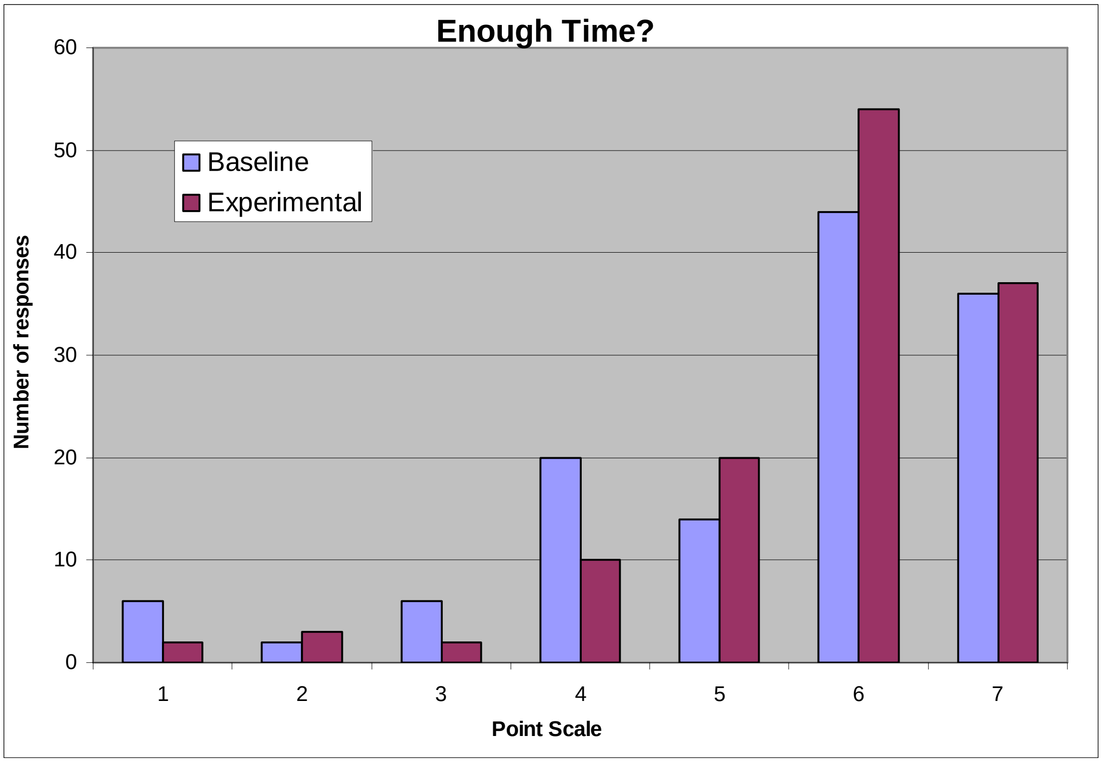


Figure 10.3 Statistics of the post-search questionnaire (Experiment II)

10.2.7 Post-system Questionnaire

Subjects' opinions about the systems were assessed by a post-system questionnaire (Appendix C(5)). Subjects were asked whether the system was easy to learn to use, easy to use, understandable and useful on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely." Table 10.11 presents the mean and standard deviation for each system for these questions. The systems were compared for each question based on each subject's responses using the non-parametric Wilcoxon signed rank test. SPSS 14.0 software was used to do the analysis.

Results showed that it was slightly, but not significantly easier to learn to use the experimental system (M=5.53, SD=1.11) than the baseline system (M=5.25, SD=1.52),

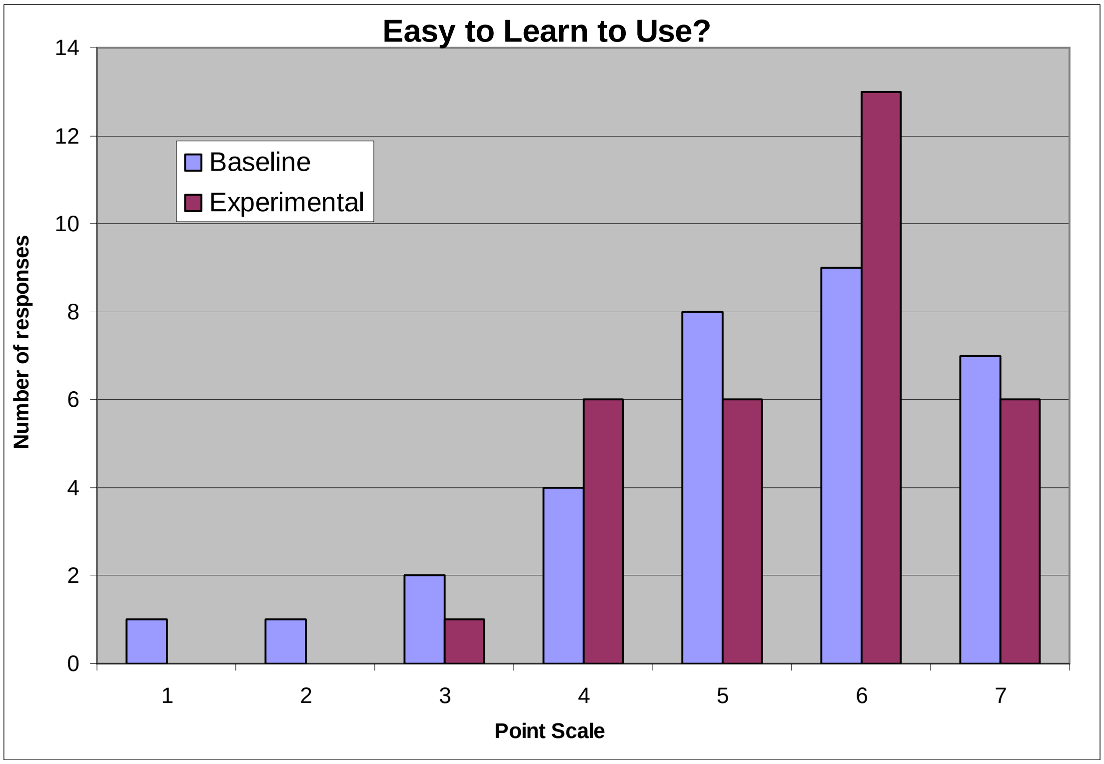
$Z=-0.744$, $p=0.457$. Subjects felt it was significantly easier to use the experimental system ($M=5.38$, $SD=1.16$) than to use the baseline system ($M=4.72$, $SD=1.33$), $Z=-2.264$, $p=0.024$. Subjects felt they understood the system better when using the experimental system ($M=5.25$, $SD=1.30$) than using the baseline system ($M=5.09$, $SD=1.49$), but the difference was not significant, $Z=-0.488$, $p=0.625$. Subjects felt the experimental system ($M=5.44$, $SD=1.32$) was significantly more useful than the baseline system ($M=4.47$, $SD=1.34$), $Z=-2.522$, $p=0.012$. Figure 10.4 shows the distribution of responses to the post-system questionnaire.

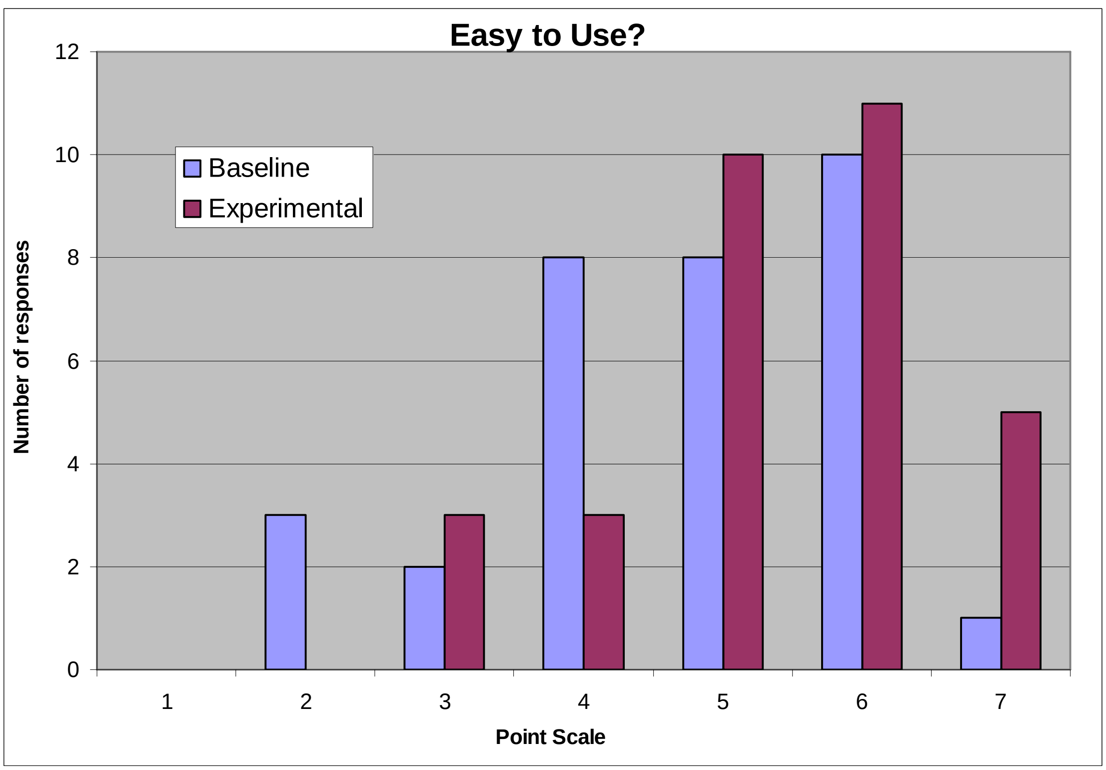
Table 10.11

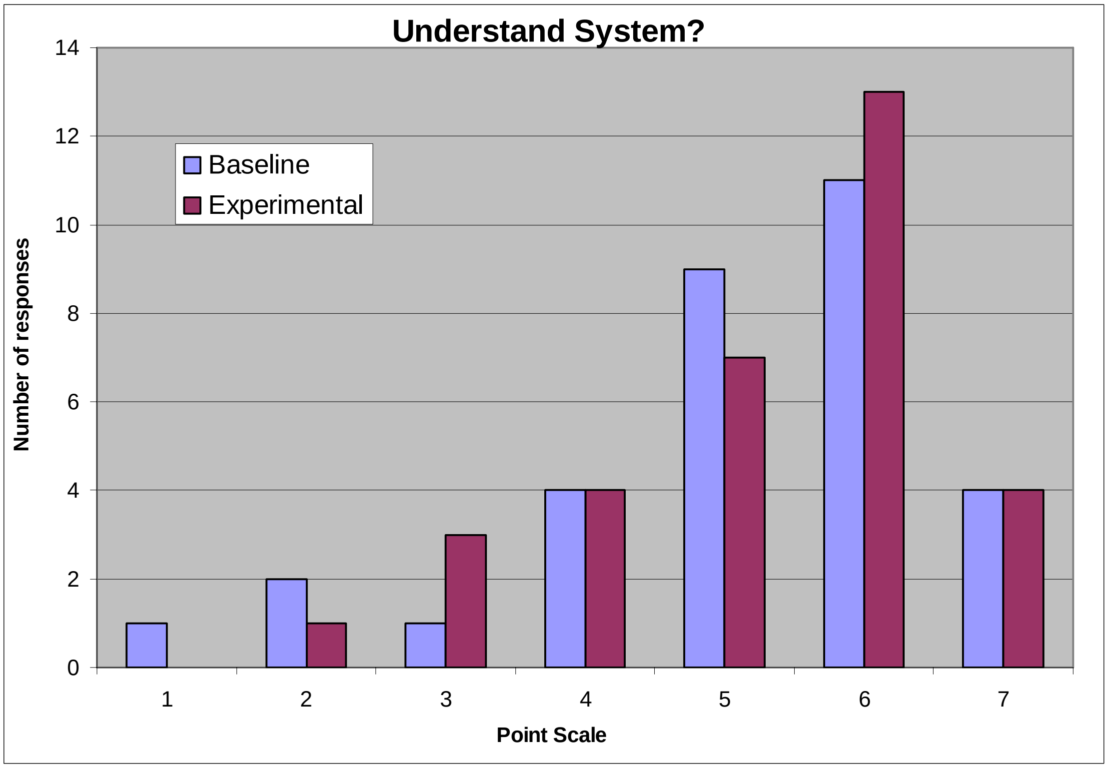
Post-system Questionnaire Results (Experiment II)

	Systems	Easy to learn to use?	Easy to use?	Understand system?	Usefulness of system?
Mean	Baseline	5.25	4.72	5.09	4.47
	Experimental	5.53	5.38	5.25	5.44
Standard deviation	Baseline	1.52	1.33	1.49	1.34
	Experimental	1.11	1.16	1.30	1.32
p-value	Baseline	0.457	0.024*	0.625	0.012*
	Experimental				

*$p < 0.05$


Easy to Learn to Use?
Baseline
Experimental
Number of responses
0
2
4
6
8
10
12
14
1
2
3
4
5
6
7
Point Scale


Easy to Use?
Baseline
Experimental
Number of responses
0
2
4
6
8
10
12
1
2
3
4
5
6
7
Point Scale



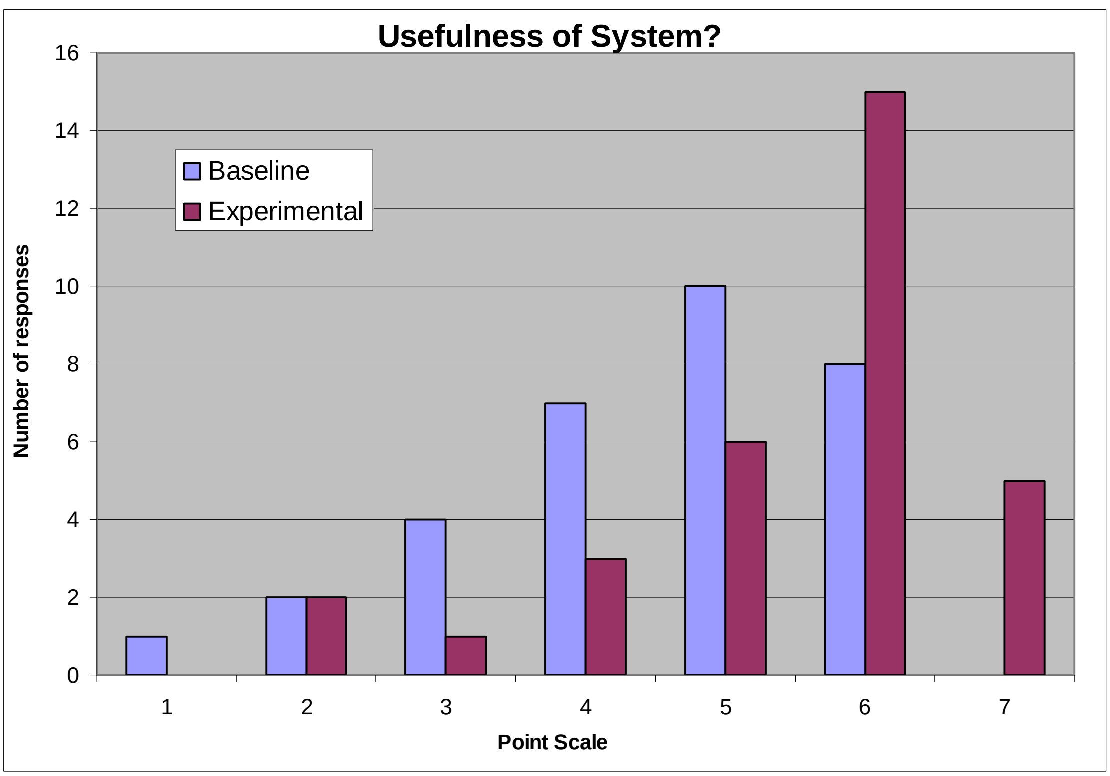


Figure 10.4 Statistics of the post-system questionnaire (Experiment II)

10.2.8 Exit Interview

An exit interview (Appendix C(6)) was presented to the subjects after they completed searching using both systems. Subjects were asked to rate how different the two systems were on a 7-point scale, where 1 = "not at all"; 4 = "somewhat"; and 7 = "extremely."

Subjects were also asked to decide which system was more helpful in completing tasks, was easier to learn to use, was easier to use, was better for the search tasks, and which system they liked best, with three choices: system 1 (either E or B), system 2 (either E or B), no difference (ND). In addition, subjects were asked which system features they liked or disliked most and were asked to give some general comments with open-ended questions. For each question, the Sign test (ignoring no difference) was employed to test whether the number of subjects who preferred the experimental system was significantly different from the number of subjects who preferred the baseline system. Subjects found E and B were different at a high level ($\underline{M}$=5.03, $\underline{SD}$=1.23).

In Table 10.12, subjects believed E was significantly more helpful than B (E=21, B=6, ND=5), $\underline{p}$=0.007. Results showed that B was easier to learn to use than E (E=6, B=16, ND=10), but the difference was not significant, $\underline{p}$=0.052. E was easier to use than B (E=18, B=8, ND=6), although not significantly so, $\underline{p}$=0.078. Results showed that E was significantly better to search than B (E=23, B=7, ND=2), $\underline{p}$=0.006. Overall, significantly more subjects liked E best (E=25, B=7, ND=0), $\underline{p}$=0.003.

Figure 10.5 shows the distribution of responses to the exit interview.

Table 10.12

System Comparison of the Exit Interview (Experiment II)

Systems	Question abstract				
	More helpful?	Easier to learn?	Easier to use?	Better for search	Best overall?
Baseline	6	16	8	7	7
Experimental	21	6	18	23	25
No difference	5	10	6	2	0
p-value	0.007*	0.052	0.078	0.006*	0.003*

*$p < 0.05$

A sample of the responses from the subjects to the exit interview questions is shown in Table 10.13. All quotations are exactly what the subjects said, including misspellings and other errors.

Table 10.13

Open-ended Questions (Experiment II)

Exit interview questions	Baseline system	Experimental system
Difference between two systems	"…just has pages, pages of quotes…"; "…just has a whole long list of searches…"; "…more simpler, but not necessary more helpful…"; "...much more simplistic. Has far fewer search queries, and did not suggest search terms…"	"user has much more control over the sources to search, more advanced.."; "…it enables you to view thing in more digestible amount.."; "…allowed you to search some additional clues, additional criteria…"; "…it could give me some criteria, for publisher, publish city, topic, title…"
More helpful in completing tasks	"…I actually like ….better. Even though it just has a list, and the query did not allow to put publication, etc, it was less frustrating, I found out I had to do less clicking."	"…for an electronic book, …is definitely more helpful; For tasks of searching news articles, there are least difference..."; "...reduces the labor of the researcher, and let the computer do more work…"; "…the way it allows to browse,

		like showing the book chapters. That is useful…"; "...It gave me more options to narrow my search…"
Easier to learn to use	"...it has less options…"; "..so simple, you can just do something if you want to do something…"; "…close to my early experience, for example, Google"; "… no learning about it. Just type in…"	"…I need to see what is going to happen and I need to spend sometime to learn if they are going to be beneficial to my search…"; "… it took me a while to figure out what is going on…"
Easier to use	"… you just write the topic you want to search for.."; "...is kind of hard, because you have to change the search constantly to find the correct information…"; "… less clicking. There are fewer initial options. I could just jump on to it…"	"it gives the descriptions about the databases… cluster results… some suggestion box…"; "..less amount of data to choose from, because of the clusters…"; "…It gives me more options…"; "… When you go into the book, it expands the chapters for you…"
Able to conduct better search?	"…very difficult to read. Part of the time consumption was used for reading the headlines…"; "…require me to browse through as many documents…"; "… if you only have 10 minutes, the simpler one is better…"; "…too many irrelevant hits…"	"…more detailed information for input, more accurate results…"; "..it limits the output and help you get what you want…."; "… The setup of suggested topics are right there, just click and go…"; "… able to look around within the book…"
Best overall?	"…just quote…"; "retrieving everything related to my word, but also retrieved some that has nothing to do with the word… precision was not good…"; "… fewer intermediate screens…"	"…don't care how complicated the system is. The best thing is we can get good results…"; "… more organized…"; "…always think of the benefit and the handiness that it provides for book search…"; "… put more emphasis on what I am able to give to rather than how I can get things back…"; "… it reminds me of online catalog,… it organizes the articles in clusters…"
Features that you like most?	"…both systems, I like that there is a division between news articles and books…"; "…list of paragraphs in a single book…"; "…the	"…table of contents. Search of author. Cluster results…"; "… highlight of the words…"; "… Database description…"; "…clusters give you general

	paragraphs summarizing the books…”	picture of the search results in terms of quantity…”; “…The list of suggested topics are great…”; “…Pre-selected topics help narrow down things…”; “…I like the high level idea of outputting results in hierarchy… like clusters…”; “…It still gave me the name that approximates my spelling…”
Features that you dislike most?	“…results for the book, .. just give me the whole thing, not just some paragraphs…”; “…just one section, it is hard to do, because it look like it is close to getting actual data, but just isolated, can not scroll up or down…”; “…the massive list … hard to read…”; “…can not do a search in the content text…”	“…the clusters… slows down the process…the simpler the better”
Features to add?	“…search …within the results of previous search…”; “…Quick find, or ctrl+F feature…”; “…a mouse copy and paste, rather than using keyboard…”; “…A drop down menu for search history…”; “…Emailing and printing…”; “…percentage relevance…”; “…Logical queries to both…”	

Table 10.14 shows a list of features that the subjects liked, in the descending order of number of subjects that liked the features.

Table 10.14

Features that Subjects Liked (Experiment II)

Features	No. of subjects that liked the feature	Systems
Field searches	21	E
Clusters	14	E
Table of contents	10	E
Database description	9	E
Highlighting	6	E & B
Suggested topics	4	E
Spelling check	2	E & B
Suggestion box	1	E
Logical searching (and/or)	1	E

Note: E = Experimental; B = Baseline.

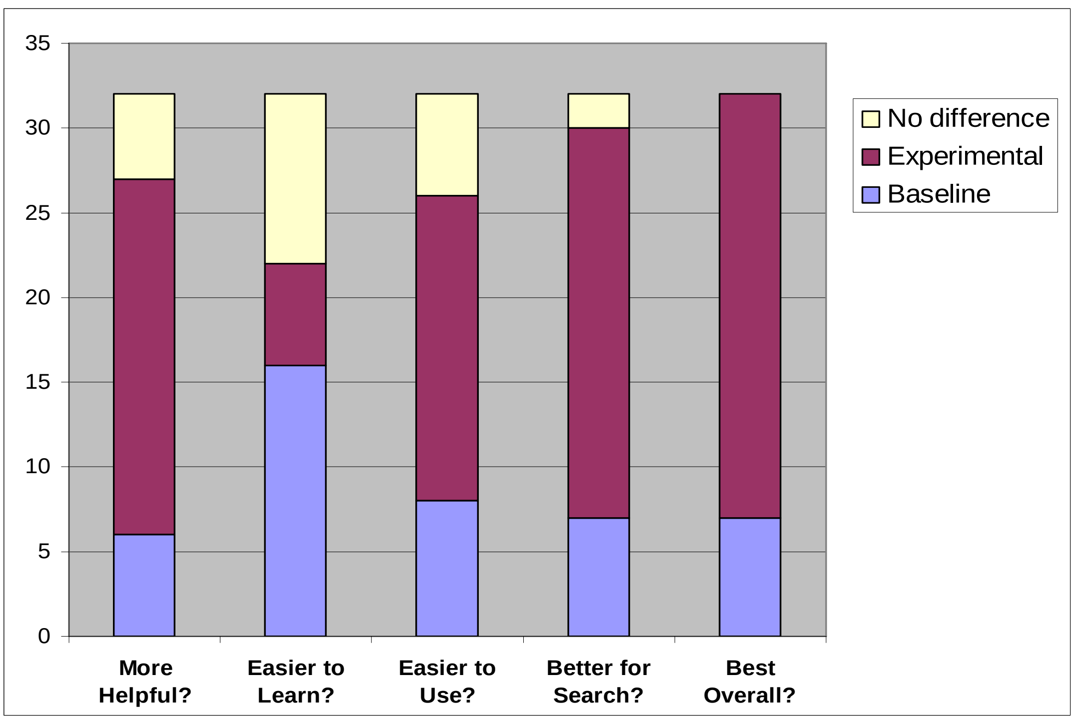


Figure 10.5 Statistics of the exit interview (Experiment II)

Chapter 11

Research Problem 3: Discussion

This chapter discusses the results of the second experiment from three perspectives: performance, interaction, and usability.

The measures on which the experimental system significantly outperformed the baseline system are listed in Table 11.1. There were no cases in which the baseline system significantly (or even non-significantly) out performed the experimental system.

Table 11.1

Measures with Significant Results Favoring the Experimental System (Experiment II)

Categories		Measures	p-value
Performance		Aspectual recall	0.009
		Result satisfied	0.008
Interaction		Number of iterations	0.035
		Query length	0.000
Usability	Post-search questionnaire	Easy getting started	0.025
		Enough time	0.014
	Post-system questionnaire	Easy to use	0.024
		Usefulness of system	0.012
	Exit interview	More helpful	0.007
		Better for search	0.006
		Best overall	0.003

The results of our experiments demonstrated that the experimental system adapted to different ISSs within the course of a single information-seeking episode had significant advantages over the baseline system which was designed to support specified searching only. There was no measure on which the baseline system, built using current state-of-the-art technology, and using the standard current IR support techniques, outperformed the experimental system. Furthermore, in each of our evaluation categories, the experimental system significantly outperformed the baseline system, on at least two measures. Hypothesis 5, that the experimental system designed for supporting both scanning and searching performs better in supporting integrated tasks requiring both scanning and searching than the baseline system designed for supporting integrated tasks with one specific method, was confirmed in terms of performance, interaction and usability measures.

11.1 Performance

With respect to our performance measures, although there was no significant difference between the systems with respect to time taken for completing the task, this was probably an artifact of the design, which limited search time to twelve minutes. Some subjects mentioned in the exit interview that they could do better searches if more time was given. The exit interview result also indicated that since the experimental system provided more options and features, it took subjects more time to explore. The baseline system may have taken less time because it only provided the ranked list of documents or paragraphs, which was similar to their most familiar search engine, Google.

There was no significant difference in correctness of results between the systems. However, subjects using the experimental system found more correct answers and fewer wrong answers than those using the baseline system (see Table 10.6). This might be explained from the exit interview results that subjects believed that the experimental system offered them “more control” and options on the input query (see Table 10.13). The non-significant result could be attributed to the limited time of the experiment itself.

Based on ANOVA test, there were no significant differences in the effect of system order or task order on time and result correctness. There were also no significant differences in the interaction between system order and task order on time and result correctness.

The advantage to the experimental system for both aspectual recall and user satisfaction with results was highly significant. This indicated that the experimental system helped subjects identify more relevant aspects from the given documents in multiple databases; and subjects felt more satisfied with results they found from the experimental system. The exit interview results gave a good explanation for this. Subjects commented in the exit interview that the features provided in the experimental system such as clustering and fielded query gave them flexibility to "narrow down" their queries, table of contents provided them a "general picture" of the whole book, the results were more "organized" using clusters, and the descriptions of the databases made it easier to choose the right database to search (see Table 10.13).

11.2 Interaction

There were significant differences in favor of the experimental system on two interaction measures.

Subjects using the experimental system had significantly fewer iterations than those using the baseline system. Subjects using the experimental system employed significantly longer queries than those using the baseline system. Since it was known that longer queries performed better in best match systems, the latter result was of some general interest. Further analysis showed that the queries used in the experimental system were still significantly longer than the queries used in the baseline system, even if fielded queries were excluded in the analysis.

11.3 Usability

Results from post-search and post-system questionnaires, and also the exit interview, strongly support that the experimental system was more usable than the baseline system.

Subjects felt it was significantly easier to start their tasks in the experimental system, and when asked if they had sufficient time to do the search, they gave significantly more positive responses for the experimental system. It was interesting to notice that the subjects spent more time using the experimental system than using the baseline system. This conflict between the objective measure of time and the subjective measure of time may be because that the subjects could get more accurate answers from the experimental system and they were more satisfied with the results, thus it made them think they had more enough time for the experimental system. Since there were a lot more features in the experimental system, the subjects may also take the time to explore different features, which could prolong the search time.

Both the experimental and the baseline systems, in many respects, were novel to the subjects. But subjects felt the experimental system was significantly easier to use and more useful than the baseline system, with respect to the tasks. In the exit interview, subjects mentioned that the experimental system was more like an advanced system which provided more options and control to help them narrow down their search and get more relevant results.

Subjects felt it was easier to learn to use the experimental system than to learn to use the baseline system although no significant result was found. Ratings of this sort might be subjected to the problem of “demand characteristic” which is defined as “a term used in psychology experiments to describe a cue that makes subjects aware of what the experimenter expects to find or how subjects are expected to behave. Demand characteristics can change the outcome of an experiment because subjects will often change their behavior to conform to the experimenters expectations" (http://psychology.about.com/od/dindex/g/demanchar.htm). Subjects felt they understood the experimental system better than the baseline system but not significantly so. This could be attributed to the fact that all subjects favored the Google search engine which provided ranked list of retrieval results.

Subjects felt the experimental system was significantly more helpful and helped them conduct significantly better searches than the baseline system. In the exit interview, subjects identified many features they liked in the experimental system, such as table of contents, fielded query, database selection and topic lists. They claimed that all these features made the experimental system more helpful and helped them conduct better searches.

The result on which system was easier to use from the post-system questionnaire was not consistent with the result from that of the exit interview (see Table 11.2). This might be caused by the inconsistent rating of the subjects.

Table 11.2

Comparison of the Post-system Questionnaire and the Exit Interview (Experiment II)

Questionnaire / Interview	Systems	Easier to Learn?	Easier to Use?
Post-system	Baseline (B)	0.457 (E)	0.024* (E)
	Experimental (E)		
Exit	Baseline	0.052 (B)	0.078 (E)
	Experimental		

*$p < 0.05$

It was also considered whether the results could have arisen from any systematic differences between the subjects with respect to their topic expertise or familiarity. The data on these factors (see Table 10.9) did not seem to support this, as the subjects' mean self-reported expertise and familiarity, measured on a 7-point low to high scale, were uniformly low for all topics, with rather low standard deviation as well.

Thus, it appears that our basic hypothesis, that a system which adapts to support different ISSs during the course of an information-seeking episode performs better than a system designed to support only the standard ISS of specified searching, is supported.

Chapter 12

Conclusions

The overall goal of the study is to construct and evaluate an IIR system which supports a searcher engaging in a variety of different ISSs in different ways during the course of an information-seeking episode. This system is based on a theoretical model of IIR which construes an information-seeking episode as a person's moving from one ISS to another and a classification of ISSs.

Four IR support techniques (database summary, table of contents navigation, clustered retrieval results, and fielded query) were identified to best support the different ISSs that a searcher might engage in while attempting to resolve different kinds of information problems. These techniques were incorporated into four experimental systems each of which was compared to a respective baseline system in a within-subjects experiment (experiment I). Results showed that systems tailored to different IR support techniques can better support different ISSs for different varieties of tasks, than generic IR systems designed to support specified searching. It was also shown that systems incorporating these techniques are more usable than the respective baseline systems. Thus, it could be concluded from experiment I that an IR system should be designed by incorporating different IR support techniques for different ISSs and tasks.

In order to better provide guidance for the interaction between the user and the system, some pre-existing dialogue structures were adopted to implement an experimental (integrated) system which adapted to support both scanning and searching behaviors within a single framework.

To see whether this experimental system would in fact better support human information-seeking than the baseline system (the type of IIR system designed to support only one kind of ISS: specified searching comparing a query to a set of information objects), a within-subjects experiment (experiment II) was conducted. This experiment compared user performance and behavior in our experimental system to that in a baseline system which emulated the support offered by most standard IIR systems. The experimental system is based on an explicit model of IIR which attempts to relate various characteristics of the user in the system, including the user's context, to different ISSs in which the person might engage in, and to relate the different ISSs to one another in a systematic way.

The results of experiment II demonstrated substantial and significant advantage of the experimental system in terms of objective and subjective performance (measured as aspectual recall and result satisfaction); degree of user interaction with the system (measured as number of iterations and mean query length); and usability (measured as ease of getting started, enough time, ease of use, usefulness, helpfulness and ability to conduct better searches). These results speak strongly in favor of the general concept of designing IIR systems explicitly to support different ISSs. They also demonstrate that it is possible to support quite different behaviors within a single system framework which searchers can understand and use effectively. They also demonstrate that a principled approach to designing such systems is possible. It has been shown through this study, at least to a limited extent, that a model of IIR as support for interaction with information, combined with an empirically-based classification of such interactions, can provide such principles.

There are limitations to the conclusions which can be drawn from this study, and there are issues which can be further investigated.

Firstly, as always in user studies of this type, the experiments were constrained by a small and to some extent homogeneous sample of subjects, and also by a small number of search topics. The only realistic way to address this issue is to do more studies, which we intend to perform.

Secondly, since this was an experimental study, the subjects were assigned topics to search, rather than searching on topics of their own interest, and searched in somewhat limited databases. This problem was addressed by using scenario-based topic descriptions (Borlund, 2000), and by use of a TREC collection, but the only way to really deal with it is to move from a strictly experimental environment to a quasi-experimental environment in which the experimental system is embedded in a real-life context. Such a study awaits a more robust and complete system than the one that has been tested, as well as one that is not so specifically tailored to particular types of information problems.

Thirdly, the experimental system was tailored to only a small number of different ISSs. The identification of robust IR support techniques for other ISSs, their implementation in a more general integrated IIR system, and the evaluation of such a system is an obvious next step.

Fourthly, the subjects recruited were all graduate students and had above-average searching experience. This limits the generalization of the results to other groups, such as novices and experts. Conducting more studies by extending the sampling to such groups should be another fruitful direction.

Fifthly, the patterns of ISSs in different tasks should be identified. A Markov model could be constructed in order to identify the appropriate patterns of ISSs in a variety of tasks.

Finally, a longitudinal study could be considered to complement the current experimental design which would address limited time, and limited and non-realistic tasks. Such a study would be most productive if conducted when more ISSs and IR support techniques have been tested and incorporated in the integrated system.

Despite the limitations and unanswered questions associated with this study, it is an important step on the road toward adaptive IIR systems. This research contributes several major findings. Firstly, it showed that the effectiveness of IIR systems could be improved by providing explicit and principled connections among varieties of ISSs, different IR support techniques and IIR system design. Secondly, it constructed and evaluated an integrated and novel IIR system. Finally, it suggested new methods of structuring interaction in IIR.

Appendix A

A Sample Topic from HARD 2004 Corpus

<topic>
<number>
HARD-428
</number>
<title>
International organ traffickers
</title>
<description>
Who creates the demands in the international ring of organ trafficking?
</description>
<topic-narrative>
Many countries are institutionalizing legal measures to prevent the
selling and buying of human organs. Who, in the ring of international
organ trafficking, are the "buyers" of human organs? Any information
that identifies 'where' they are or 'who' they may be will be
considered on topic; the specificity of info does not matter. Also,
the story must be about international trafficking. Stories that only
contain information about the "sellers" of organs or those that focus
on national trafficking will be off topic.
</topic-narrative>
<metadata-narrative>
Subject (CURRENT EVENTS) is chosen as it is expected that such
articles will have more information about the identities of the
parties involved. Genre (NEWS) is expected to exclude stories that
tends to focus on ethical matters.
</metadata-narrative>
<retrieval-element>
passage
</retrieval-element>
<metadata>
<familiarity>
little
</familiarity>
<genre>
news-report
</genre>
<geography>
any
</geography>
<related-text>
<on-topic>

Every day, 17 Americans die of organ failure. In Israel, the average wait for a kidney transplant is four years. In response, a global gray market has bloomed. In India, for example, poor sellers are quickly...
</on-topic>
<relevant>
At least 30 Brazilians have sold their kidneys to an international human organ trafficking ring for transplants performed in South Africa, with Israel providing most of the funding, says a legislative...
</relevant>
</related-text>
<subject>
CURRENT EVENTS
</subject>
</metadata>
</topic>

Appendix B(1)

Consent Form (Experiment I)

Searcher #: ____________
Searcher Name: ____________
Time/Date: ____________________

Thank you for volunteering to participate in this study, which aims to study how different combinations of support techniques can be used to effectively support different information-seeking strategies in different situations and tasks.

This study will be conducted in Rutgers. Your participation will entail engaging in the following activities, which will take about one hour and a half:

1. You will read and sign this consent form and ask any questions that you may have. You will receive a copy of this two-part consent form for your future reference. This should take about 5 minutes.
2. You will fill out a questionnaire about your background, computer experience and previous searching experience, which should take about 5 minutes.
3. You will fill out a pre-search questionnaire before you start each search. This should take about 2 minutes.
4. You will be given four different search tasks on which you will perform information searches using two information retrieval systems. You will be given up to 10 minutes to conduct each search. The interaction between you and the system will be logged by the system.
5. After completing each search, you will be asked to complete a post-search questionnaire. This should take about 2 minutes for each search.
6. After you finished two searches per system, you will be asked to complete a post-system questionnaire. This should take about 2 minutes for each system.
7. You will be asked to fill out an exit questionnaire after you have completed searching so that we can learn more about your experience with the systems. This should take about 5 minutes.

The results of your searches will be reported, but without any reference to you specifically. The names of all searchers will be held confidential, and all results will be reported anonymously. The analyses of questionnaires and the log of your searches will be cumulated with those of all of the other searchers for reporting purposes, and when analyzed individually will be done without reference to specific individuals, thus insuring anonymity.

The data that are collected will be used for research purposes. Names will not be attached to the log and questionnaires, and these data will be available only to the researchers on this project. Unless you explicitly agree to allow further use of these data, they will all remain confidential and will be destroyed on completion of the research study.

There are no foreseeable risks to participation in this study. You may feel pressured or nervous due to the test-like nature of the experimental task. Please remember that there are no "right" or "wrong" answers. Participation is voluntary. You may discontinue participation at any time without penalty.

Your participation in this study will advance the cause of information science and give you genuine research experience. You may indicate your wish to receive a copy of the written study report.

I, ____, have read and understood this description of the research study and agree to participate.

Participant Signature	Date
Investigator Signature	Date

Consent To Use Data in Future Research

Searcher #: ____________
Searcher Name: ____________
Time/Date: ____________________

I would like to ask your permission to use the data collected in this study for further research, for demonstration in teaching, and for presentation during conferences. If you do not want to give your permission for me to use your data, you may still participate in this study. If you do not want to give your permission for me to use your data, I will destroy your data as soon as I have analyzed it for the current study. Use of your data could entail any of the following:

1. Researchers, both at Rutgers and at other institutions, re-analyzing your questionnaires and the log of your searches for a future study. Such use would be only on approval of this project principal investigator.
2. Showing excerpts of the log and questionnaires during presentation of the research results of this project at scholarly conferences.

Please remember that once you have completed your participation in the study, all links between your name and your data will be destroyed. Thus, all results from the study will be reported and reanalyzed anonymously.

If you agree to our making use of your data, please sign this form in the space below. If you do not wish to permit such use, do not sign this form. In this case, the logs will be treated as previously described.

I, ____,have read and understood this description of how my data might be used in future research by the investigator and grant the investigator permission to use my data in the conditions described above.

____ ____________________________
Participant Signature Date

____ ____________________________
Investigator Signature Date

If you have any concerns or require further information, please contact Xiaojun Yuan (Principal Research Investigator) at (732) 429 4689 or via e-mail at yuanxj@rci.rutgers.edu

You may also contact Nicholas J. Belkin, who serves as Chair to this Dissertation, at 732.932.7500 ext. 8271 or nick@belkin.rutgers.edu.

If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University (732) 932-0150 ext 2104.

Participant's initials (if not signed) _________

Appendix B(2)

Entry Questionnaire (Experiment I)

Background Information

1. What undergraduate or graduate degree(s) have you earned or do you expect to earn? Please list major(s).

____
Degree Major

____
Degree Major

____
Degree Major

____
Degree Major

2. What is your gender?

☐ Female
☐ Male

3. What is your age?

☐ 16 – 25 years
☐ 26 – 35 years
☐ 36 – 45 years
☐ 46 – 55 years
☐ 56 – 65 years
☐ 66 years +

Computer and Searching Experience

1. How often do you use computer in your daily life?

Never			***Monthly***			***Daily***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. How do you rate your level of expertise with computers?

Novice						***Expert***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. Please circle the number that most closely describes your searching experience.

How much experience have you had…	***None***			***Some***			***A great deal***
a. searching on computerized library catalogs either locally (e.g., your library) or remotely (e.g., Library of Congress)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
b. searching on commercial online systems (e.g., Factiva, Dialog)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
c. searching on World Wide Web search engines (e.g., Google, AltaVista, Yahoo!, Teoma)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
d. searching on other systems (please specify): ______________________	*1*	*2*	*3*	*4*	*5*	*6*	*7*

e. How often do you conduct a search on any kind of system?	***Never***			***Monthly***			***Daily***
	1	*2*	*3*	*4*	*5*	*6*	*7*

f. When I search for information, I can usually find what I am looking for.	***Rarely***			***Some-times***			***Often***
	1	*2*	*3*	*4*	*5*	*6*	*7*

4. How do you rate you level of expertise with searching?

Novice						***Expert***
1	*2*	*3*	*4*	*5*	*6*	*7*

5. How many years have you been doing online searching? _______ years.

6. Please list your favorite search engine(s): _________________________.

Appendix B(3)

Pre-search Questionnaire (Experiment I)

(Document Task)

Topic: As a graduate student, you are asked to write an essay about global warming for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of documents on a variety of topics. You believe it would be interesting to discover factors that affect global warming, but you have no idea which databases are good on this topic.

Task: Please find out which databases are good for this particular topic, and rank the databases in order of likelihood of being good. Put your answer in the given space.

1. Please indicate how familiar you are with this topic:

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. Please indicate your level of expertise with this topic:

Novice						***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. If you think that you know any factors, please write them in the space below:

If you have answered this question, please circle the number that indicates how certain you are of these factors.

Extremely Uncertain			***Neutral***			***Extremely Certain***
1	*2*	*3*	*4*	*5*	*6*	*7*

(Book Task)

Topic: You are in the process of preparing an address on childhood education. There are a lot of books available on this topic. But what you are interested in is the history of censorship of books for kids. You recall that some comments from an electronic book named "Report of the Special Committee on Moral Delinquency in Children and Adolescents The Mazengarb Report (1954)" might be very useful for this talk. The comments talked about what kinds of publications children should not read. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Find the relevant comments from this book. Copy the related paragraphs then paste them to the given space.

1. Please indicate how familiar you are with this topic:

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. Please indicate your level of expertise with this topic:

Novice						***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. If you think that you know any comments, please write them in the space below:

If you have answered this question, please circle the number that indicates how certain you are of these comments.

Extremely Uncertain			***Neutral***			***Extremely Certain***
1	*2*	*3*	*4*	*5*	*6*	*7*

Appendix B(4)

Post-Search Questionnaire (Experiment I)

Please answer the following questions, as they relate to this specific task.

	Not at all			***Some-what***			***Extremely***
1. Was it easy to get started on this search?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
2. Was it easy to do the search on this topic?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
3. Are you satisfied with your results?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
4. Did you have enough time?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

	None			***Some***			***A great deal***
5. Did your previous knowledge of the topic help you?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
6. Have you learned anything new about the topic?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

Appendix B(5)

Post-System Questionnaire (Experiment I)

Please answer the following questions as they relate to the search experience that you just had with the information system.

	Not at all			***Some-what***			***Extremely***
1. How easy was it to learn to use this information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
2. How easy was it to use this information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
3. How well did you understand how to use the information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
4. How useful was the information system in helping you accomplish your search tasks?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

Appendix B(6)

Exit Questionnaire (Experiment I)

To have a better understanding of your overall experiences, I would like to ask you a few questions about your experiences today.

1. How different did you find the systems from one another?

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

Please specify reasons:

2. Which system is more helpful in completing tasks?

☐ System 1 ☐ System 2 ☐ No difference

Please specify reasons:

3. Which system did you find easier to learn to use?

☐ System 1 ☐ System 2 ☐ No difference

Please specify reasons:

4. Which system did you find easier to use?

☐ System 1 ☐ System 2 ☐ No difference

Please specify reasons:

5. Which system did you like best overall?

☐ System 1 ☐ System 2 ☐ No difference

Please specify reasons:

6. What system features did you like most?

7. What system features did you dislike most?

8. Do you have any other comments or suggestions?

Appendix C(1)

Consent Form (Experiment II)

Searcher #: ____________
Searcher Name: ____________
Investigator Name: ____________
Time/Date: ____________________

Supporting Multiple Information-Seeking Strategies in a Single System Framework (Experiment II)

Thank you for volunteering to participate in this study, whose goal is the development of information retrieval systems more sensitive to the intentions and behaviors of their users.

This study will be conducted in Rutgers. Your participation will entail engaging in the following activities, which will take about three hours:

1. You will read and sign this consent form and ask any questions that you may have. You will receive a copy of this two-part consent form for your future reference. This should take about 5 minutes.
2. You will fill out a questionnaire about your background, computer experience and previous searching experience, which should take about 3 minutes.
3. You will fill out a pre-search questionnaire before you start each search. This should take about 2 minutes.
4. You will perform searches using two information retrieval systems. For each system, you will be given a training task to get familiar with the system. After the training, you will be given four different search tasks to perform information searches using the same system. You will be given up to 12 minutes to conduct each search. After completing all the searches using the first system, you will get a 3-minute break before you proceed to the second system. The interaction between you and the system will be logged by the computer.
5. You are encouraged to "think aloud" about what you are doing and why you are doing it during your searches. What you say will be recorded.
6. After completing each search, you will be asked to complete a post-search questionnaire. This should take about 2 minutes for each search.
7. After you finish all the searches using one given system, you will be asked to complete a post-system questionnaire. This should take about 2 minutes.
8. You will be interviewed after you have completed searching so that we can learn more about your experience with the systems. The interview will last about 12 minutes and will be recorded.

Participant's initials (if not signed) _________

The results of your searches will be reported, but without any reference to you specifically. Your name will be held confidential, and all results will be reported anonymously. The analyses of your data will be cumulated with those of all of the other participants for reporting purposes. As soon as you have completed your participation, all links between your name and your data will be destroyed, thus insuring anonymity.

The data that are collected will be used for research purposes. Names will not be attached to the collected data, and these data will be available only to the researchers on this project. Unless you explicitly agree to allow further use of these data, as described below, they will all remain confidential and will be destroyed on completion of the research study.

There are no foreseeable risks to participate in this study. Your refusal to participate will involve no penalty. If you decide to participate in this study, you will get $30 cash equivalent value (gift card/cash) after you complete the experiment. You may discontinue participation at any time. In such case, you won't get anything.

Your participation in this study will advance the development of information science and give you genuine research experience. You may indicate your wish to receive a copy of the written study report.

I, ____, have read and understood this description of the research study including the audio-recording of my thinking aloud and the interview, and agree to participate.

Participant Signature	Date
Investigator Signature	Date

If you have any concerns or require further information, please contact Xiaojun Yuan (Principal Research Investigator) at (646) 705 4329 or via e-mail at yuanxj@rci.rutgers.edu

You may also contact Nicholas J. Belkin, who serves as Chair to this Dissertation, at 732.932.7500 ext. 8271 or nick@belkin.rutgers.edu.

If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University (732) 932-0150 ext 2104.

Participant's initials (if not signed) _________

Consent To Use Data in Future Research

Searcher #: ____________
Searcher Name: ____________
Investigator Name: ____________
Time/Date: ____________________

Supporting Multiple Information-Seeking Strategies in a Single System Framework (Experiment II)

I would like to ask your permission to use the data collected in this study for further research, for demonstration in teaching, and for presentation during conferences. If you do not want to give your permission for me to use your data, you may still participate in this study. If you do not want to give your permission for me to use your data, I will destroy your data as soon as I have analyzed it for the current study. Use of your data could entail any of the following:

1. Researchers, both at Rutgers and at other institutions, re-analyzing your data for a future study. Such use would be only on approval of the principal investigator of this study.
2. Showing excerpts of your data during presentation of the research results of this study at scholarly conferences.

Please remember that as soon as your participation ends in the study, all links between your name and your data will be destroyed. Thus, all results from the study will be reported and reanalyzed anonymously.

If you agree to our making use of your data, please sign this form in the space below. If you do not wish to permit such use, do not sign this form. In this case, the data will be treated as previously described.

I, ____, have read and understood this description of how my data might be used in future research by the investigator and grant the investigator permission to use my data in the conditions described above.

______________________________________ ___________________________
Participant Signature Date

______________________________________ ___________________________
Investigator Signature Date

If you have any concerns or require further information, please contact Xiaojun Yuan (Principal Research Investigator) at (646) 705 4329 or via e-mail at yuanxj@rci.rutgers.edu

Participant's initials (if not signed) __________

You may also contact Nicholas J. Belkin, who serves as Chair to this Dissertation, at 732.932.7500 ext. 8271 or nick@belkin.rutgers.edu.

If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University (732) 932-0150 ext 2104.

Participant's initials (if not signed) _________

Appendix C(2)

Entry Questionnaire (Experiment II)

Background Information

1. What undergraduate or graduate degree(s) have you earned or do you expect to earn? Please list major(s).

____

Degree Major

____

Degree Major

____

Degree Major

____

Degree Major

2. What is your gender?

☐ Female
☐ Male

3. What is your occupation?

____

4. What is your age? _______years

Computer and Searching Experience

1. How often do you use computer in your daily life?

Never			***Monthly***			***Daily***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. How do you rate your level of expertise with computers?

Novice						***Expert***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. Please circle the number that most closely describes your searching experience.

How much experience have you had searching for information using…	***None***			***Some***			***A great deal***
a. computerized library catalogs either locally (e.g., your library) or remotely (e.g., Library of Congress)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
b. commercial online systems (e.g., Dialog, Lexis-Nexis)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
c. World Wide Web search engines (e.g., Google, AltaVista, Yahoo!)	*1*	*2*	*3*	*4*	*5*	*6*	*7*
d. other systems (please specify): ______________________	*1*	*2*	*3*	*4*	*5*	*6*	*7*

e. How often do you conduct a search on any kind of system?	***Never***			***Monthly***			***Daily***
	1	*2*	*3*	*4*	*5*	*6*	*7*

f. When I search for information, I can usually find what I am looking for.	***Rarely***			***Some-times***			***Often***
	1	*2*	*3*	*4*	*5*	*6*	*7*

4. How do you rate you level of expertise with searching for information?

Novice						*Expert*
1	*2*	*3*	*4*	*5*	*6*	*7*

5. How many years have you been doing online searching? _______ years.

6. Please list your favorite search engine(s): ________________________.

Appendix C(3)

Pre-search Questionnaire ((Experiment II)

(News Article Task)

Topic: As a graduate student, you are asked to write an essay about global warming for one of your courses. You are supposed to get information you need from a system that is composed of several databases. Each database has lots of news articles on a variety of topics, but you have no idea which databases are good on this topic. You believe it would be interesting to discover factors that affect global warming, and would like to collect news articles that identify different factors.

Task: Please find as many different factors as possible. For each factor, please copy the title or link of the article which discusses that factor, and paste it to the answer box. For each article that you copy, please type or copy the factor(s) that it identifies. If an article discusses more than one factor, you only need to copy and paste the article once. If there are several articles which discuss the same factors, you only need to copy and paste one such article.

1. Please indicate how familiar you are with this topic:

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. Please indicate your level of expertise with this topic:

Novice						***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. If you think that you know any factors, please write them in the space below:

If you have answered this question, please circle the number that indicates how certain you are of these factors.

Extremely Uncertain			***Neutral***			***Extremely Certain***
1	*2*	*3*	*4*	*5*	*6*	*7*

(Book Task)

Topic: You are in the process of preparing a talk on the development of airplanes. There are a lot of books available on this topic. But what you are interested in are experiments which significantly affected the development of model airplanes. You recall that some comments from an electronic book might be very useful for the talk. You cannot remember the exact name of the book. But you remember that it is written by Charles Vivan, or someone like that, and was published in the early 20th century. The comments are about the first model of an airplane invented in the seventeenth century. You cannot remember the exact comments, but would like to quote them in your talk.

Task: Please find the relevant comments from the book, copy the one best paragraph then paste it into the answer box. Also, please copy the title of the book then paste it to the answer box.

1. Please indicate how familiar you are with this topic:

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

2. Please indicate your level of expertise with this topic:

Novice						***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

3. If you think that you know any comments, please write them in the space below:

If you have answered this question, please circle the number that indicates how certain you are of these comments.

Extremely Uncertain			***Neutral***			***Extremely Certain***
1	*2*	*3*	*4*	*5*	*6*	*7*

Appendix C(4)

Post-Search Questionnaire (Experiment II)

Please answer the following questions, as they relate to this specific task.

	Not at all			***Some-what***			***Extremely***
1. Was it easy to get started on this search?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
2. Was it easy to do the search on this topic?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
3. Are you satisfied with your results?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
4. Did you have enough time?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

	None			***Some***			***A great deal***
5. Did your previous knowledge of the topic help you?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
6. Have you learned anything new about the topic?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

Appendix C(5)

Post-System Questionnaire (Experiment II)

Please answer the following questions as they relate to the search experience that you just had with the information system.

	Not at all			***Some-what***			***Extremely***
1. How easy was it to learn to use this information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
2. How easy was it to use this information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
3. How well did you understand how to use the information system?	*1*	*2*	*3*	*4*	*5*	*6*	*7*
4. How useful was the information system in helping you accomplish your search tasks?	*1*	*2*	*3*	*4*	*5*	*6*	*7*

APPENDIX C(6)

Exit Interview (Experiment II)

To have a better understanding of your overall experiences, I would like to ask you a few questions about your experiences today.

1. How different did you find the systems from one another?

Not at all			***Somewhat***			***Extremely***
1	*2*	*3*	*4*	*5*	*6*	*7*

Why?

2. Which system is more helpful in completing tasks?

☐ System 1 ☐ System 2 ☐ No difference

Why?

3. Which system did you find easier to learn to use?

☐ System 1 ☐ System 2 ☐ No difference

Why?

4. Which system did you find easier to use?

☐ System 1 ☐ System 2 ☐ No difference

Why?

5. In which system were you able to conduct better searches?

☐ System 1 ☐ System 2 ☐ No difference

Why?

6. Which system did you like best overall?

☐ System 1 ☐ System 2 ☐ No difference

Why?

7. What system features did you like most? Why?

8. What system features did you dislike most? Why?

9. What other system features would you suggest to be added to the systems? Why?

10. Do you have any other comments or suggestions?

References

Allan, J. (2005). HARD Track overview in TREC 2004: High accuracy retrieval from documents. In E.M. Voorhees & L.P. Buckland (Eds.), *TREC 2004, Proceedings of the Thirteenth Text Retrieval Conference*. Washington, DC: Government Printing Office.

Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*(4), 205-214.

Bates, M. (1990). Where should the person stop and the information search interface start? *Information Processing & Management, 26*(5), 575-591.

Belkin, N. J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. In *Information Retrieval '93*, Germany, 55-66.

Belkin, N. J. (1996). Intelligent information retrieval: Whose intelligence? *In Proceedings of the Fifth International Symposium for Information Science* (ISI-96), 25-31.

Belkin, N. J., Chang, S., Downs, T., Saracevic, T., & Zhao, S. (1990). Taking account of user tasks, goals and behavior for the design of online public access catalogs. In *Proceedings of the 53rd ASIS annual meeting,* Leaned Information, Inc., Medford, NJ., 69-79.

Belkin, N. J., Cool, C., Jeng, J., Keller, A., Kelly, D. Kim, J., Lee, H.-J., Tang, M.-C., & Yuan, X.-J. (2002). Rutgers' TREC 2001 Interactive Track Experience. In E.M. Voorhees & D.M. Harman (Eds.), *TREC 2001, Proceedings of the Tenth Text Retrieval Conference* (pp. 465-472). Washington, D.C.: Government Printing Office.

Belkin, N. J., Cool, C., Kelly, D., Lin, S-J., Park, S. Y., Perez-Carballo., J., & Sikora, C. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing & Management, 37*(3), 403-434.

Belkin, N. J., Cool, C., Koennman, J., Ng, K. B., Park, S. (1996). Using relevance feedback and ranking in interactive searching. In E.M. Voorhees & D.M. Harman (Eds.), *TREC-4, Proceedings of the Fourth Text Retrieval Conference* (pp. 181-209). Washington, D.C.: Government Printing Office.

Belkin, N. J., Cool, C., Kelly, D., Lee, H.-J., Muresan, G., Tang, M.-C., & Yuan, X.-J. (2003). Query length in interactive information retrieval. In *Proceedings of the 26th Annual ACM International Conference on Research and Development in Information Retrieval (SIGIR '03),* Toronto, CA, 205-212.

Belkin, N. J., Cool, C., Stein, A., & 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.

Belkin, N. J., Marchetti, P. G., & Cool, C. (1993). BRAQUE: Design of an interface to support user interaction in information retrieval. *Information Processing & Management, 29*(3), 325-344.

Borlund, P. & Ingwerson, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation, 53*(3), 225-250.

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

Chang, S. & Rice, R.E. (1993). Browsing: A Multidimensional Framework. *Annual Review of Information Science and Technology, 28,* 231-278. Edited by Williams, M.E.

Chang, S. (1995). *Toward a multidimensional framework for understanding browsing.* Unpublished Ph.D. dissertation, SCILS, Rutgers University, New Brunswick, NJ.

Chen, H. & Dhar, V. (1991). Cognitive processes as a basis for intelligent retrieval system design. *Information Processing & Management, 27*, 405-432.

Cool, C. & Belkin, N. J. (2002). A classification of interactions with information. *Proceedings of the Fourth International Conference on Conceptions of Library and Information Science,* 1-15.

Croft, W. B. & Thompson, R. H. (1987). I^3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science, 38*(6), 389-404.

Cutrell, E., Robbins, D. C., Dumais, S. T., & Sarin, R. (2006). Fast, Flexible Filtering with Phlat – Personal Search and Organization Made Easy. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006),* Canada, 261-270.

Cutting, D. R., Karger, D. R., Pedersen, J. O., & Turkey, J. W. (1992). Scatter/gather: A cluster-based approach to browsing large document collections. *In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'92).* 318-329.

Dumais, S. & Belkin, N. J. (2005). The TREC interactive tracks: Putting the user into search. In E.M. Voorhees & D.K. Harman (Eds.) *TREC: Experiment and Evaluation in Information Retrieval* (pp. 123-152). Cambridge, MA: MIT Press.

Egan, D. E., Remde, J. R., Gomez, L. M., Landauer, T. K., Eberhardt, J., & Lochbaum, C. C. (1989). Formative design-evaluation of SuperBook. *ACM Transactions on Information Systems, 7*(1), 30-57.

Ellis, D. (1989). A behavioral approach to information retrieval design. *Journal of Documentation, 45*(3), 171-212.

Ellis, D. D. Cox, and Hall, K. (1993). A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of Documentation, 49*(4), 356-369.

Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1). 61-74.

Frisse, M. & Cousins, S. B. (1989). Informaiton retrieval from hypertext: Update on the dynamic medical handbook project. In hypertext '89 Proceedings. New York, ACM. 199-212.

Harter, S. P. (1986). *Online information retrieval: Concepts, principles, and techniques.* Orlando, FL: Academic Press.

Harter, S. & Hert, C. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. In M.Williams(Ed.)*, Annual review of information science and technology, 32,* 3-94.

Harter, S. P. & Rogers-Peters, A. (1985). Heuristics for online information retrieval: A typology and preliminary listing. *Online Review, 9*(5), 407-424.

Hawkins, D. T., & Wagers, R. (1982). Online bibliographic search strategy development. *Online, 6*(3), 12-19.

Hearst, M. (1995). Tilebars: Visualization of term distribution information in full text information access. *In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, CO., USA, 59-66.

Hearst, M. , Pedersen, J., Pirolli, P., Schuze, H., Grefenstette, G., & Hull, D. (1996). Xerox Site Report: Four TREC-4 Tracks. In D. Harman (ed.) *TREC-4, Proceedings of the Fourth Text Retrieval Conference.* Washington, DC: Government Printing Office.

Ingwerson, P. (1992). *Information Retrieval Interaction.* London, UK: Taylor Graham.

Ingerwersen, P. (1996). Cognitive Perspectives of Information Retrieval Interaction. *Journal of Documentation, 52*(1), 3-50.

Jardine, N. & van Rijsbergen, C. J. (1971). The use of hierarchic clustering in information retrieval. *Information Storage and Retrieval, 7,* 217–240.

Kuhlthau, C. C. (1991). Inside the search process: information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*, 361-371.

Lin, S.-J. (2001). *Modeling and supporting multiple information seeking episodes over the web.* Unpublished Ph.D. dissertation, SCILS, Rutgers University, New Brunswick, NJ.

Markey, K., & Atherton, P. (1978). ONTAP: on-line training and practice manual for ERIC database searchers. ERIC Clearinghouse on Information Resources, Syracuse University.

Marchionini, G. (1995). Cambridge series on human-computer interaction: *Information seeking in electronic environments.* Cambridge: Cambridge University Press.

Marchionini, G., Brunk, B. (2006). Toward a general relation browser: A GUI for information architects. Retrieved on May 25, 2007, from http://www.science.oas.org/Ministerial/ingles/documentos/marchionini_general_relation.pdf

Meadow, C.T. & Cochrane, P. (1981). *Basics of online searching*. New York: Wiley.

Muresan, G. (2002). *Using Document Clustering and Language Modelling in Mediated Information Retrieval.* PhD thesis, School of Computing, Robert Gordon University, Aberdeen, Scotland, United Kingdom.

Oddy, R.N. (1977). Information retrieval through man-machine dialogue. *Journal of Documentation, 33* (1), 1-14.

Olston, C. & Chi, Ed H. (2003). ScentTrails: Integrating browsing and searching on the web. *ACM Tarnsactions on Computer-Human Interaction, 10*(3), 177-197.

Over, P. (2001). The TREC interactive track: an annotated bibliography. *Information Processing & Management, 37*, 369-381.

Pejtersen, A.M. (1989). A library system for information retrieval based on a cognitive task analysis and supported by an icon-based interface. *Proceedings of the 12th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Cambridge, MA., 40-47.

Reichman, R. (1985). *Getting computers to talk like you and me.* Cambridge, MA: MIT Press.

Saracevic, T. (1995). Evaluation of evaluation in information retrieval. *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Special issue of SIGIR Forum, 138-146.

Saracevic, T. (1996). Modeling interaction in information retrieval (IR): A review and proposal. *Proceedings of the American Society for Information Science, 33,* 3-9.

Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. *Proceedings of the American Society for Information Science, 34,* 313-327.

Shute, S.J. and Smith, P.J. (1993). Knowledge-based search tactics. *Information Processing & Management, 29* (1), pp. 29-45.

Sitter, S., & Stein, A. (1992). Modeling the Illocutionary Aspects of Information-Seeking Dialogues. *Information Processing and Management, 28* (2), 165-180.

Spink, A. (1997). Study of Interactive Feedback during Mediated Information Retrieval. *Journal of the American Society for Information Science, 48* (5), 382-394.

Stein, A. & Thiel, U. (1993). A Conversational Model of Multimodal Interaction. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI '93),* Menlo Park, CA: AAAI Press/ The MIT Press, 283-288.

Strohman, T., Metzler, D., Turtle, H. and Croft, W. B. (2005). "Indri: A language-model based search engine for complex queries (extended version)," (CIIR Technical Report), IR-407, CIIR, University of Massachusetts.

Thompson, R. & Croft, W. B. (1989). Support for browsing in an intelligent text retrieval system. *International Journal of Man-Machine Studies, 30*, 639-668.

Winograd, T. & Flores, F. (1986). *Understanding computers and cognition.* Norwood, NJ: Ablex.

White, R. W., Ruthven, I., & Jose, J. M. (2003). Finding relevant documents using top ranking sentences: An evaluation of two alternative schemes. *Proceedings of the 24th Annual International Conference on Research and Development in Information Retrieval (SIGIR'03),* Tampere, Finland, 57-64.

Xie, H. (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science, 51*(9), 841-857.

Yee, P., Swearingen, K., Li, K. and Hearst, M. (2003). Faceted metadata for image search and browsing. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2003),* Florida, USA, 401-408.

VITA

Xiaojun Yuan

Education

2000-2007 Rutgers University, New Brunswick, NJ
Ph.D. in Communication, Information and Library Studies, October 2007
M.S. in Statistics, January 2003
1997-2000 Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
Ph.D. in Computer Application, July 2000
1991-1997 Xi'an University of Science & Technology, Xian, China
M.E. in Computer Application, July 1997
B.E. in Computer Application, July 1994

Publications

Yuan, X.-J., & Belkin, N.J. (2007). Supporting multiple information strategies in a single system framework. In: C.L.A. Clarke, N. Fuhr, N. Kando, W. Kraaij, A.P. de Vries (Eds.) In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR'07)* (pp. 247-254). New York: ACM.

Murdock, V., Kelly, D., Croft, W.B, Belkin, N.J., **Yuan, X.-J.** (2007). Improving retrieval for task questions with question-independent structural features. *Journal of Information Processing and Management. 43,* 181-203.

Zhang, X.-M., Anghelescu, H., **Yuan, X.-J.** (2005). Domain knowledge, search behavior, and search effectiveness of engineering/science students: An exploratory study. *Information Research: An International Electronic Journal.* 10(2), paper 217 (2005). [Available at http://InformationR.net/ir/10-2/paper217.html]

Belkin, N.J., Cool,C., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., Tang, M.-C., **Yuan, X.-J.** (2003). Query length in interactive information retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '03),* Toronto, Canada, 205-212.

Wacholder, N., Sharp, M., Liu, L, **Yuan, X.-J.**, Peng Song (2003). Experimental study of index terms and information access. In *Proceedings of the Annual Conference of the American Society for Information Science and Technology (ASIST '03),* Long Beach, CA, 184-192.

Kelly, D., Murdock, V., **Yuan, X.-J.**, Croft, W.B. & Belkin, N.J. (2002). Features of documents relevant to task- and fact-oriented questions. In *Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM '02)*, McLean, VA, 645-647.

Yuan, X.-J., Belkin, N.J., Kim, J. (2002). The relationship between ASK and relevance criteria. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '02),* Tampere, Finland, 359-360.

Belkin, N.J., Cole, M. , Gwizdka, J. , Li, Y.-L. , Liu, J.-J., Muresan, G., Roussinov, D. , Smith, C. L., Taylor, A., **Yuan, X.-J.** (2006). Rutgers Information Interaction Lab at TREC 2005: Trying HARD. In D. Harman & E. Voorhees (Eds.), *TREC2004, Proceedings of the Fourteenth Text Retrieval Conference*. Washington, D.C.: GPO.

Belkin, N. J., Kelly, D., Lee, H.-J., Li, Y.-L., Muresan, G., Tang, M.-C., **Yuan, X.-J.,** and Zhang, X.-M. (2005). Rutgers' HARD Experiences at TREC 2004. In D. Harman & E. Voorhees (Eds.), *TREC2004, Proceedings of the Thirteenth Text Retrieval Conference*. Washington, D.C.: GPO.

Belkin, N. J., Kelly, D., Lee, H.-J., Li, Y.-L., Muresan, G., Tang, M.-C., **Yuan, X.-J.,** and Zhang, X.-M. (2004). Rutgers' HARD and Web Interactive Track Experiences at TREC 2003. In D. Harman & E. Voorhees (Eds.), *TREC2003, Proceedings of the Twelveth Text Retrieval Conference*. Washington, D.C.: GPO.

Belkin, N. J., Cool, C., Kelly, D., Kim, G., Kim, J., Lee, H.-J., Muresan, G., Tang, M.-C., & **Yuan, X.-J.** (2003). Rutgers' TREC 2002 Interactive Track Experience. In D. Harman & E. Voorhees (Eds.), *TREC2002, Proceedings of the Eleventh Text Retrieval Conference*. Washington, D.C.: GPO.

Belkin, N. J., Cool, C., Jeng, J., Keller, A., Kelly, D., Kim, J., Lee, H.-J., Tang, M.-C., & **Yuan, X.-J.** (2002). Rutgers' TREC 2001 Interactive Track Experience. In D. Harman & E. Voorhees (Eds.), *TREC2001, Proceedings of the Tenth Text Retrieval Conference*. Washington, D.C.: GPO.