Assessment of the Effects of User Characteristics on Mental Models of Information Retrieval Systems

Xiangmin Zhang
Library and Information Science Program, Wayne State University, Detroit, MI 48202.,
E-mail: ae9101@wayne.edu

Mark Chignell
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada.
E-mail: chignel@mie.utoronto.ca

This article reports the results of a study that investigated effects of four user characteristics on users' mental models of information retrieval systems: *educational and professional status*, *first language*, *academic background*, and *computer experience*. The repertory grid technique was used in the study. Using this method, important components of information retrieval systems were represented by nine concepts, based on four IR experts' judgments. Users' mental models were represented by factor scores that were derived from users' matrices of concept ratings on different attributes of the concepts. The study found that *educational and professional status*, *academic background*, and *computer experience* had significant effects in differentiating users on their factor scores. *First language* had a borderline effect, but the effect was not significant enough at $\alpha = 0.05$ level. Specific different views regarding IR systems among different groups of users are described and discussed. Implications of the study for information science and IR system designs are suggested.

Introduction

User diversity calls for information retrieval (IR) systems that can accommodate heterogeneous user groups (Allen, 1996). To design such systems, it is necessary to identify and understand different types of users.

Efforts on user-centered research in IR have been made for decades. Individual difference studies in IR have found that user search performance varied on certain user characteristics (Bellardo, 1985; Borgman, 1989; Charoenkitkarn, 1996; Fenichel, 1981; Kamala, 1991; Marchionini et al., 1993; Qiu, 1993; Woelfl, 1984; Yee, 1993). Such characteristics include a user's experience with a system, academic background, age, gender, and personality (Borgman, 1989; Egan, 1988). For example, in terms of academic background, Borgman (1984b) and Kamala (1991) found science/engineering majors had better search performance than social sciences/humanities majors did.

In general, this type of studies found that search performance differences exist on certain user characteristics and the findings are helpful for understanding users, to some extent. The major limitation of these studies, however, is that the cognitive processes or the reasons why one type of users would perform better than or differently from another type of users were not studied. Further investigations are needed to better understand the performance differences as well as how to effectively design IR systems that accommodate these differences (Savage-Knepshield & Belkin, 1999). For system designs, we not only need to know on what characteristics users' behavior or search performance would vary, but also need to know why: Why different types of users behave in different ways and have different search performance.

Research in IR interaction emphasizes searching as an interactive task and the user's interaction with IR systems. In addition to the comparisons between different system/interface designs for revealing the nature of interactions and system/user search performance (Voorhees & Harman, 1998), a number of interactive IR models have been proposed, such as Belkin's (Belkin & Vickery,1985; Belkin et al., 1995) episode model, Ingwersen's (1992) cognitive model, and Saracevic's (Saracevic, 1997; Spink & Saracevic, 1997) stratified model. These models not only reflect the interactive nature of IR systems but also try to explain, from different perspectives, at the cognitive level the reasoning processes behind user's interaction with IR systems.

While providing frameworks for studying users in the process of directly consulting an IR system (Robins, 2000),

Received February 22, 1999; Revised July 6, 2000; accepted July 25, 2000.

these models generally do not address individual difference issues. All users are treated as equal, not distinguishing different types. In addition, these models lack details. It is not clear how users, particularly different types of users, would interact with IR systems. Empirical research is needed which could provide details on how and why people's interaction with IR systems would differ and what are the differences in people's understanding of IR systems.

Users' search behavior or performance can be further understood by exploring their mental models. There have been a number of mental model studies in the field of IR (Borgman, 1986; Cool et al., 1996; Dimitroff, 1992; Kerr, 1990). Most of these studies investigated the relationships between mental models and users' search performance. For example, Borgman (1986) compared user performance and methods/patterns of searches of model-based training with that of nonmodel-based training on an on-line catalog, assuming different training methods would generated different mental models. Her study found that both the performance and the search patterns were different on complex tasks between model-based and nonmodel users. Dimitroff (1992) studied relationships between mental models and search performance. Subjects were categorized into four groups based on the completeness of their mental models. The results of the study show that subjects with more complete models made significantly fewer errors and found significantly more items.

These studies, however, did not compare or reveal differences of mental models between user groups that can be easily characterized. The models were either on individual basis, which can hardly be generalized, or were categorized on some criterion that are as hard to identify as model themselves do. The findings are thus offer little help for designing systems that can accommodate users with different characteristics. In addition, the methodological issues in these studies raised concerns about the "models" obtained. For example, Borgman (1986) pointed out the methodological difficulty of capturing a person's mental model in her study.

In summary, user-centered research in IR so far has generated a rich knowledge about users and users' interaction with IR systems. Unfortunately, there has not been much research on the different views or mental models between user groups with different characteristics. How different types of users view a system and thus interact with it differently is still rarely known. For systems design, it is difficult to assign users to models or stereotypes based on differences in their cognitive abilities (Allen, 2000).

This article reports the results of a study on users' mental models of IR. Different from the research reviewed above, this study meant to differentiate mental models held by different types of users and to associate user characteristics with the models. The exploration of the differences of mental models of IR systems is important because variance of user performance will be explained by the model differences. Individual differences among users of IR systems can thus be better understood. It is also helpful for designs of IR systems. Knowing what users think about IR can form the basis of tailoring user characteristics-related features that accommodate particular user groups.

Mental models generally are the models people have of the things with which they interact (Norman, 1990). For this study, a mental model is defined as the user's understanding of the components and their interconnections of a computer system, and the processes that change the components (Borgman, 1986; Carroll & Olson, 1988).

The reason to study mental models is that user's interaction with a given computer system is guided by the user's mental model of that system (Borgman, 1986; Carrol & Olsen, 1988; Halasz, 1983; Kieras & Polson, 1985; Norman, 1990; Preece et al., 1994; Sebrechts et al. 1990). These models enable users to predict system performance and form the basis of the interactions with the system (Faulkner, 1998). By studying mental models, users' anticipations or views towards a system could be revealed and be used in designs of IR systems to accommodate different types of users.

Studying mental models poses open methodological issues. The fundamental difficulty in this area is that, compared with other models, such as a physical or a mathematical model, a mental one is not observable. Users cannot show it explicitly. They normally have difficulty giving an adequate account of the structure and content of their own knowledge (Briggs, 1987; Norman, 1983). This feature makes it very difficult to elicit and to measure a mental model, not to mention identifying the differences between models. "Even if we agree that models are good, we're not sure how to distinguish a good model from a bad one . . ." (Borgman, 1984a).

There are a number of methods for determining mental models. This study was based on Kelly's (1955) Repertory Grid Technique (RGT). In the remainder of this article, the RGT, as well as the reasons for using it in this study, is briefly introduced and discussed in Repertory Grid Technique. Research questions and hypotheses established for this study are then described in Research Questions. The Research Design section includes a description of variables and the experimental design based on the RGT. The results of data analyses are presented in the Results section. The findings are discussed in the Discussion section, which is followed by Implications, in which the implications of the study for both information science research and IR system designs are discussed. The article ends with Conclusions and Future Research.

Repertory Grid Technique (RGT)

The RGT was invented as a tool for and is based on Kelly's personal construct theory (Kelly, 1955). The core of Kelly's theory is that people understand the world (events, people, etc.) through their personal construct systems. The personal construct system that each person develops is the set of representations or model of the world that the person has developed. It is acquired through the person's social experience.

The RGT, in its simplest form, involves the generation of a list of concepts (elements) about things or events to be investigated, and the generation of attributes (constructs) based on the list of concepts.

A concept is defined as anything that can be compared or contrasted. For example, people, vegetables, or notions such as occupations, feelings, situations, events, etc., can all be elements. If the problem is to choose a future career, the concepts may be different jobs. Concepts used in a study may be elicited from the subject or provided by the tester, or both, and they need to be well known and personally meaningful to the subject (Shaw, 1980).

An attribute (construct) is a bipolar dimension that, to some degree, is a property of each concept. A construct is a way in which some things (elements) are seen as alike yet still different from others. 'It is essentially a two-ended affair, involving a particular basis for considering likenesses and differences of elements and at the same time for excluding certain things as irrelevant to the contrast involved" (Bannister, 1968). Examples of attributes for concepts about people may be: *Don't believe in God/Very religious; Not athletic/Athletic; Understands me better/Doesn't understand at all; Sociable/Not sociable,* etc. (Fransella & Bannister, 1977).

Like concepts, attributes can also be elicited from the subjects or provided by the tester. There are several ways to elicit attributes. The classic method used by Kelly is to consider various triads (groups of three concepts) selected successively from the whole concept list. The subject or person(s) from which attributes are to be elicited is first presented with three concepts and asked to specify some important aspects in which two of them are alike. Then the subject is asked in which aspects the third concept differs from the other two. Often the subject will indicate spontaneously which two concepts are being judged alike. The subject's description of the similarity forms one pole of the attribute and the answer to the question concerning the difference is the contrast pole. Such a process is called a sort. The examiner records this similarity and contrast as the resulting attribute dimension from the first sort, and proceeds to the second and subsequent sorts using different triads of concepts. There are no rules on how many triads of concepts should be presented to the subjects, but between 10 and 25 is a common range (Bannister, 1968; Fransella & Bannister, 1977).

Presently, the most frequently used variation of repertory grids is rating grids. In a rating grid, the subject is asked to evaluate the concepts systematically by using the attribute list to generate the grid of rating numbers. At every intersection of column and row is the subject's rating value of the concepts on the attributes. The grid form was used as both a model elicitation tool and as a formal representation of mental models in this study. This matrix of numbers (grid) can be looked upon as a map of an individual's mental model of the system or situation to be investigated.

For a detailed description of the repertory grid technique please refer to Fransella and Bannister (1977).

Owing to its power in exploring a person's unspoken views on given things, the RGT has been used independently as a research tool in many fields other than psychology. These include education, management, knowledge engineering, and mental models of information retrieval systems. A thorough review of the technique's applications can be found in Gaines and Shaw (1991).

The meaningfulness of the personal construct theory to mental models research lies in that it identifies individual constructions of experience as the source of a person's behavior. It could be "a means to evaluate knowledge. More specifically it could be used as a tool for the assessment of the user's conceptual model and therefore to identify those aspects of a system which are most commonly misunderstood" (Briggs, 1987). Latta and Swigger (1992) study has showed the applicability of the method in modeling users of IR systems.

The RGT was chosen for the study because it has several advantages for eliciting, representing, and analyzing mental models over other methods such as thinking aloud, interview, subject observation, etc. First, subjects do not have to use verbal descriptions to describe their mental models, which is hard for some people. Second, the grid form provides a unified representation for all subjects and this formal representation makes it much easier to compare between different models; and third, represented by numeric grid, the technique makes it possible that mental models can be statistically analyzed and compared. Unstructured representations such as verbal protocols make comparisons of mental models difficult. These advantages served well for this study's purpose of investigating differences in mental models held by different types of users. The way the method was used is described in the Research Design section.

Research Questions

The general purpose of this study was to associate user characteristics with mental model features. Specifically, the study intended to explore the effects of four user characteristics on users' mental models of IR systems: *educational and professional status*, *first language*, *academic discipline*, and *computer experience*. Educational and professional status was the educational degree level if a user was studying at school or the working status if the person was employed by the time of this study. The reason for choosing this characteristic was that it normally reflects a person's knowledge and skills in a professional field. Because information retrieval is a highly intellectual activity, it is natural to believe that educational and professional status has an impact on mental models.

A user's first language is the language the user acquired at home during his or her childhood. It is also called the user's native language. This characteristic was chosen because IR systems are closely related to languages.

Academic discipline or background refers to the major area of knowledge a user was studying. Different disciplines have different bodies of knowledge and different approaches to exploring knowledge. These differences may have an impact on users' mental models.

Computer experience is referred to as a subject's experience in using any of a list of computer applications such as database management, electronic mail, information retrieval, etc. Because IR systems are presumably computerized systems, a person's computer experience was considered important in shaping the person's views of IR systems.

This study sought to answer the following five research questions: (a) did a user's educational and professional status have an effect on the user's mental model of IR systems? (2) Did a user's first language have an effect on the user's mental model of IR systems? (3) Did a user's academic background have an effect on the user's mental model of IR systems? (4) Did a user's computer experience have an effect on the user's mental model of IR systems? (5) If there would be an effect for a characteristic, what would be the difference(s) between different types of users in their mental models?

The first four questions were then transformed into four (null) hypotheses, correspondingly.

Hypothesis H0 1: subjects with different educational and professional status do not differ in their mental models.
Hypothesis H0 2: subjects with different first languages do not differ significantly in their mental models.
Hypothesis H0 3: subjects from all academic backgrounds do not differ significantly in their mental models.
Hypothesis H0 4: subjects over all levels of computer experience do not differ significantly in their mental models.

The alternative hypotheses for Hypotheses H0 (1) to H0 (4) were that there was a significant difference in mental models among and between different groups of subjects, at the $\alpha = 0.05$ level. These hypotheses assumed the relationships between two types variables: user characteristics and mental models. The hypotheses could be tested by appropriate statistical procedures. The measures of the variables and the statistical procedures for hypotheses testing are discussed in the following section: Research Design.

Question 5 was a general question. It could follow any of the first four questions for which a positive answer, i.e., a significant effect was found. The question asked for the specific differences between groups if a significant effect was found upon a user characteristic. If no significant effect was found, i.e., the hypothesis (in null hypothesis format) for that question was accepted, the answer to Question 5 for that characteristic was not further sought.

It should be noted that the study did not intend to determine what a complete mental model was of a single user or a group of users. Instead, it focused on examining differences between groups of users in their models.

TABLE 1. Concepts and attributes used in final data analysis.

Concepts	Attributes
1. browsing	1. form/process
2. classification	
3. data structure	2. targeted/untargeted
4. document content	
5. feedback	3. specific to IR systems/applicable to all information systems (ISs)
6. information need	
7. interface	
8. query	
9. search	

Research Design

Independent Variables: User Characteristics

Educational and Professional Status was represented by four categories: working librarians who had completed professional education, graduate students, undergraduate students, and high school students in grades 11–13.

First Language was treated as a binary variable: either native English or nonnative English.

Academic Discipline was mainly divided into two broad categories: one was social sciences and humanities, and the other was engineering and science. In addition, the subjects whom could not be classified into the two categories, i.e., librarians and high school students, were also included in the data analysis on this variable.

Computer Experience: based on the frequencies of using these applications, subjects were divided into three groups: high, medium, and low.

Dependent Variables: Mental Models

Based on the RGT method, mental models in this study were represented and measured by subjects' ratings on nine concepts and three attributes about IR systems. The concepts and attributes are listed in Table 1. These concepts were suggested and decided by a group of IR experts who were University Toronto faculty and doctoral students. The nine concepts cover important components of IR process, from users' information needs as start and documents from databases as the end.

The attributes were generated by the same group of experts using the triads method. Eight attributes were generated and rated by the subjects initially. However, only the listed three attributes were able to differentiate features of mental models among the subject groups. Because the purpose of the study was to find out differences between user groups, only the concept ratings against these three were involved in the final data analyses. The concept ratings on other attributes that did not distinguish subject groups were discarded. A detailed description of concepts and attributes generating and selecting can be found in Zhang (1998). How the concepts were rated by the subjects is explained later in the Data Collection section.

The three attributes provided dimensions from which the components (concepts) of IR systems could be examined. The concept ratings displayed a subject's judgements of the relationship between concepts and attributes, and this relationship "reflects psychological assumptions underlying those judgements" (Fransella & Bannister, 1977, p. 23).

One of the problems using grid data is the large number of variables for statistical data analyses. Each rating (a combination of a concept and an attribute) is a variable and the number of variables would be the product of the number of concepts and the number of attributes. To reduce the number of variables and to explore the potential variations in original ratings, these concept ratings were summarized by a factor analysis. The resulting factor scores from subjects were employed to represent mental models. The factor analysis on concept ratings will be discussed later.

Using the RGT method to elicit and represent mental models, the major research steps and data flow is described in Figure 1.

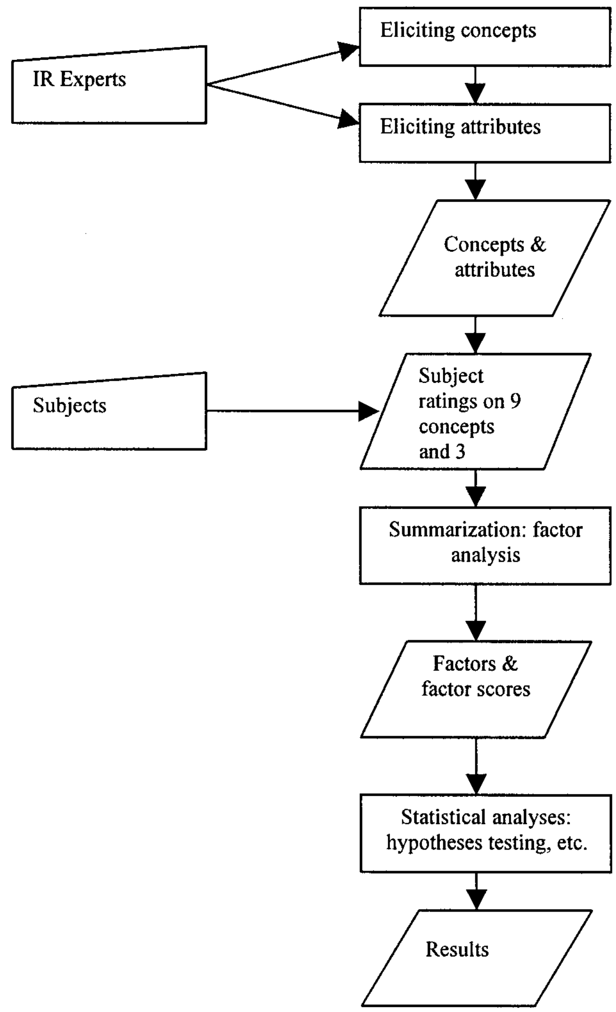

FIG. 1. Research steps and data flow.

Subjects

Sixty-four subjects voluntarily participated in this study. Subjects were recruited from four populations: professional librarians and information specialists, graduate students, undergraduate students, and high school students. Librarians and information specialists were those who had completed their professional education and qualification and were working for a library or an information agency. They might have different academic backgrounds. But after a few years of working in an information service environment, they were assumed to be very knowledgeable about IR systems, no matter in which field they had obtained their degrees.

Graduate students, at doctoral or master level, are university students who are more or less involved in research activities. Searching information for research is a frequent task for them. Therefore, they are assumed to be familiar with IR process although it is not necessary for them to know how IR systems function.

Undergraduate students are university students who are working toward their bachelor's degrees. Their main academic activity is to take courses. They may not be familiar with IR systems, but they presumably have had experience in using the university libraries.

High school students were limited to those who were at grade level 11 to 13. They were close to graduating from high school and were preparing to enter universities. They had acquired a basic pool of scientific knowledge, and were eager to learn new things.

The distribution of subjects with different educational and professional status, different languages, different disciplines, and different computer skill levels is summarized in Table 2.

In Table 2, the first column lists the numbers of subjects on educational and professional status. The second column lists the numbers of subjects according to their first language: native English and nonnative English (Non-Engl). In the third column, the distribution of the subjects in two academic disciplines is presented: Science/Engineering (abbreviated as Sci./Engi.) and Social Sciences/Humanities (abbreviated as Soc./Hum.). It should be noted that the subjects in these categories were only university students (both graduate and undergraduate). Librarians and high school students were not able to be included. The last column, column 4, presents the distribution of subjects in terms of their computing experience.

Data Collection

User Background Information

Subjects' background information about their characteristics was collected using a background questionnaire. The data collected were used for classifying subjects into different groups.

TABLE 2. Distribution of experiment subjects.

Educational and professional status	First language		Academic discipline		Computer experience		
	English	Non-Engl.	Sci. & Engi.	Soc. & Hum.	High	Med.	Low
Librarian	4	4	N/A*	N/A	8	0	0
Graduate	10	8	9	9	9	9	0
Undergraduate	4	10	6	8	7	4	3
High school	7	17	N/A**	N/A	2	9	13
Total	25	39	15	17	26	22	16

* "N/A": not applicable to librarians. They were recruited as a group who were particularly trained for information retrieval tasks, and they were not asked to state their undergraduate or graduate disciplines other than library and information science.

** The high school students did not yet have a formal academic orientation.

Concept Ratings

Subjects were asked to rate on a worksheet the nine concepts against the eight initial attributes. On the worksheet, all attributes were transformed into five point scales, with "1" at the left poles and "5" at the right poles of the attributes. In case some subjects had difficulty in understanding an attribute or a concept, or they thought an attribute was not applicable to a concept, a "not applicable" option was added to the scales, which was represented by an "X" sign. Subjects could simply circle this sign to rate a concept. A sample worksheet is shown in Figure 2.

To complete the tasks, subjects either individually met with the author in the Faculty of Information Studies, University of Toronto or they sent in their data to the author through surface or electronic mail. As mentioned earlier, only the ratings on the listed three attributes were involved in the data analyses.

Summarization of Raw Concept Ratings

Each individual rating of a concept on an attribute constituted a variable. Altogether, there were nine concepts and three attributes, constituting 27 ratings or variables. The data were summarized to reveal unexpected dimensions (or factors) among the original variables and to reduce the number of original variables to fewer ones (Mulaik, 1972).

Using the principal components approach in factor analysis, with the varimax rotation, the original 27 variables were transformed into principal factors. The first nine factors were selected to use because their eigenvalues were greater than 1, which is a norm used in factor analyses. These nine factors accounted for 68% of the total variations from the original ratings.

	Concept (1)*	
Attribute(1)(left pole)	1 2 3 4 5 X	**Attribute**(1)(right pole)
Attribute(2)(left pole)	1 2 3 4 5 X	**Attribute**(2)(right pole)
...		...
Attribute(8)(left pole)	1 2 3 4 5 X	**Attribute**(8)(right pole)

* "Concept" and "Attribute" were replaced with real ones in the study

FIG. 2. Sample concept rating worksheet.

Each of the nine factors represented certain original variables (ratings). Factor loadings and interpretations of the factors are summarized in Table 3.

The first factor can be interpreted as a dimension reflecting the intent of querying. The ratings covered by the factor were all on the attribute of *targeted/untargeted*. The concepts included were *information need*, *query*, *search*, and *document content*. These concepts are all closely related to querying or searching. Therefore, the factor was labeled Purposefulness of Querying.

The concepts summarized by the second factor are *data structure*, *document content*, *feedback*, *interface*, and *classification*. A common characteristic of these concepts may be the way of organizing and displaying data. Because these concepts were rated against the attribute of *specific to IR systems/applicable to all ISs*, and this attribute can be considered as a dimension of applicability of a component in IR systems, the factor may be viewed as a dimension reflecting the applicability of data organizations, and was labeled Applicability of Data Organization.

Factor 3 grouped together ratings of concepts related to querying on the function dimension: *form/process*. The factor therefore was labeled Function of Querying.

Factor 4 was labeled Applicability of Querying, because the querying concepts were grouped again but on the attribute of *specific to IR systems/applicable to all Iss*.

Both Factors 5 and 6 contained a mixture of different concepts and different attributes. The factors were labeled Applicability of Browsing and Function of Data Structure, respectively, based on the first rating covered in the factors.

Factors 7, 8, and 9 each had only one rating. Factor 7 was labeled Purposefulness of Browsing because the rating represented by the factor was the concept *browsing* on the attribute *targeted/untargeted*. Similarly, Factor 8 was labeled Function of Document and Factor 9 Purposefulness of Data Structure.

On each factor, a subject had a factor score. A high factor score means the concepts were rated on the high value end of the attribute scale in the factor. A low score means the

TABLE 3. Rotated factor structure for concept ratings.

Factor and variable*	Factor loadings**
Factor 1: Purposefulness of querying	
Information need: *targeted/untargeted*	0.84
Query: *targeted/untargeted*	0.82
Search: *targeted/untargeted*	0.69
Document content: *targeted/untargeted*	0.57
Factor 2: Applicability of data organization	
Data structure: *specific to IR systems/applicable to all Iss*	0.77
Doc. Content: *specific to IR systems/applicable to all Iss*	0.72
Feedback: *specific to IR systems/applicable to all Iss*	0.65
Interface: *specific to IR systems/applicable to all Iss*	0.65
Classification: *specific to IR systems/applicable to all ISs*	0.50
Factor 3: Function of querying	
Information need: *form/process*	0.73
Query: *form/process*	0.70
Search: *form/process*	0.60
Factor 4: Applicability of querying	
Query: *specific to IR systems/applicable to all Iss*	0.82
Information need: *specific to IR systems/applicable to all ISs*	0.66
Search: *specific to IR systems/applicable to all Iss*	0.53
Factor 5: Applicability of browsing	
Browsing: *specific to IR systems/applicable to all Iss*	0.85
Feedback: *form/process*	0.52
Factor 6: Function of data structure	
Data structure: *form/process*	0.74
Interface: *targeted/untargeted*	0.72
Factor 7: Purposefulness of browsing	
Browsing: *targeted/untargeted*	0.86
Factor 8: Function of document	
Document content: *form/process*	0.84
Factor 9: Purposefulness of data structure	
Data structure: *targeted/untargeted*	0.79

* Variables (ratings) within each factor are presented in descending order of their factor loadings. Each variable consists of a concept and an attribute that is italicized.

** Factor loadings are sorted by factor and only those loadings greater than 0.50 are shown.

concepts were rated on the low value end of the scale. It was hypothesized that with different educational and professional status, different languages, different disciplines, and different levels of computer experience, the factor scores would differ significantly.

Data Analysis

MANOVA, ANOVA, and Tukey tests were used together in this study to conduct statistical analyses on subjects' factor scores. For each user characteristic (independent variable), a MANOVA was conducted first to test the corresponding null hypothesis. In all MANOVA tests, the Wilk's Lambda criterion was used. If a significant difference was found, the results of ANOVA on each of the dependent variables (nine factors) were consulted to reveal on which factor(s) the difference was/were found. Tukey test, one of the multiple range tests for pairwise comparisons of means, was then employed, if necessary, to compare between groups regarding the same independent variable. Otherwise, the independent variable would be assumed no effect on factor scores and no further analysis was conducted.

Results

Results of hypothesis testing using MANOVA on hypotheses 1 to 4 are first reported. Detailed illustration of the ANOVA and Tukey testing results follows, which answers research Question 5 about those characteristics that the null hypothesis was rejected by the MONOVA test, which means that the characteristic had an effect of differentiating users' factor scores. The differences between subject groups are first described by the results of ANOVA and Tukey test. Actual average concept ratings from different groups are then compared to reveal the differences on specific concept ratings. If a characteristic was found no significant main effect by MANOVA test, no results of ANOVA and Tukey test will be presented.

Did Each of the Four User Characteristics Have an Effect on Mental Models of IR Systems? Results of MONOVA Tests

MONOVA tests were run first to test the null hypotheses that these characteristics had no effects on mental models. Each characteristic is a category to identify subjects. For

TABLE 4. Results of H0 tests on four user characteristics.

User characteristics	Tests of H0: no significant difference between groups/levels	Results of MANOVA
Educational and professional status	Rejected	$F(27,152.51) = 5.04$, $p < 0.01$
First language	Accepted	$F(9, 54) = 2.01$, $p = 0.056$
Academic disciplines	Rejected	$F(27,152.51) = 5.47$, $p < 0.01$
Computer experience	Rejected	$F(18,106) = 1.87$, $p = 0.027$

each category, subjects were classified into different groups: There were four groups based on educational and professional status: librarians, graduate students, undergraduate students, and high school students.

There were also four groups identified based on academic background: science/engineering, social sciences/humanities, professionals and "no major." The first two groups consisted of university students, both graduate and undergraduate subjects. Librarians were considered as professionals, and high school students were classified as "no major" due to their lack of clear discipline orientations. (Although librarians and high school subjects had been compared as different educational and professional status, it was reasonable to look at them from the perspective of academic background, comparing them with other disciplines.)

On first language, subjects were assigned into two groups: native English speakers and nonnative English speakers. Based on their computer experience, subjects were grouped into three levels: high, medium, and low.

The results of MANOVA tests showed that among the four characteristics, three had effects on mental models (factor scores). The test results are summarized in Table 4.

Educational and professional status, discipline, and computer experience each had a significant effect on factor scores. The null hypotheses about the three characteristics were rejected. The main effect found on educational and professional status was significant, $F(27, 152.51) = 5.04$, $p < 0.01$, meaning there was a significant difference among the different groups of subjects in their factor scores. The alternate hypothesis was supported.

A significant, $F(27, 152.51) = 5.47$, $p < 0.01$, difference was found by the MANOVA test on the characteristic of academic discipline also. The alternate hypothesis of H0 (3) was accepted: subjects from overall all academic backgrounds differed significantly in their factor scores.

On computer experience, the main effect was significant but weak, with $F(18, 106) = 1.87$ and $p = 0.027$. Difference was found in factor scores among all subjects over all levels of computer experience.

The null hypothesis on first language was accepted: there was no significant difference between the two language groups. The effect found by the MANOVA procedure was a borderline difference on their factor scores, $F(9, 54) = 2.01$, $p = 0.056$. However, the difference was not significant at the preset $\alpha = 0.05$ level. Nonnative English subjects did not differ significantly from native English ones in their factor scores. A 0.05 or a 0.01 significance level is an established norm in statistical testing. A 0.10 significance level could have had been used in this study, and that could have had accommodated the significance of the difference between the two language groups. But the 0.10 level would have too much risk for Type I errors and it was thus not used.

Were these variables interfering with each other? Was there any effect caused by the interaction of these variables? An analysis on the interactions of these variables found no effect. None of the interactions had a significant impact on factor scores, although three of the four variables did have individually. Because of unbalanced subject allocations as pointed out earlier in this article, undergraduate science/engineering had only nonnative English subjects, and librarians and high school subjects had no academic divisions, and the interaction of first language and academic background was not included in the analyses of variable interactions.

What would be the Difference(s) between Different Groups of Users? Results of ANOVA and Tukey Tests

Effects of Educational and Professional Status on Factor Scores. The results of ANOVA and Tukey test on educational and professional status are summarized in Table 5.

Among nine factors, ANOVA found that the differences occurred on factors 1, 2, 3, and 6. Table 5 (and the tables for other characteristics) displays results on only the factors that there are differences found.

As exhibited in Table 5, when different status of education and profession were compared in pairs, significant differences were found between librarians and students on factors 1 and 3. On factor 2, there was a significant difference between graduate and high school subjects. The difference between graduate and undergraduate subjects was found on factor 6. All the differences were significant at p $\leq$ *0.05 level.*

No significant difference was found between undergraduate and high school subjects.

Differences between Librarians and Students The major findings of the study are the differences between librarians and students. To clearly display the differences between different groups, every group's average factor scores and

TABLE 5. Results of ANOVA and Tukey test on factors for educational and professional status.

Status compared between:	Difference found on: f_1^*(F = 10.22, df = 3, Pr = 0.0001**)	f_2(F = 3.49, df = 3, Pr = 0.021**)	f_3(F = 11.02, df = 3, Pr = 0.0001**)	f_6(F = 2.99, df = 3, Pr = 0.038**)
Librarian and graduate	Yes***		Yes	
Librarian and undergraduate	Yes		Yes	
Librarian and high school	Yes		Yes	
Graduate and undergraduate				Yes
Graduate and high school		Yes		
Undergraduate and high school				

* f_1 to f_6 stand for Factor 1 to Factor 6. Only the factors where a difference was found are listed. Refer to Table 3 for interpretations of these factors.
** Results of ANOVA test. Type III sum of squares was used.
*** A "Yes" means a difference, significant at α = 0.05 level, was found between the two groups on the factor.

their average ratings on the concepts represented by a factor are listed in Table 6. Students' scores contrasted sharply with the librarians' scores on factor 1. While librarians tended to get very high scores on the factor, in contrast, all student groups rated querying concepts on the low value end of *targeted/untargeted* scale. The difference indicated that when considering querying factor, librarians thought of it untargeted, *while all student subjects considered it to be targeted*.

Students' scores also drastically differed from that of librarians on factor 3, which covered ratings of querying concepts on the attribute of *form/process*. While librarians' scores were very low, indicating they thought querying absolutely as *form*, students on the other hand, assigned high values on the factor, tending to consider querying as *process*.

Differences between Student Groups The factor scores from the student groups differed between themselves. Both undergraduate and high school subjects' scores contrasted to the graduates' one on factor 2, which included ratings of data organization concepts on *specific to IR systems/applicable to all ISs* dimension. Graduate subjects had a high score, indicating data organization applicable to all ISs. Both undergraduate and high school ones had a low score on this factor. However, only high school subjects were found significantly different from graduate subjects. They seemed to consider data organization applicable only to IR systems.

Undergraduate subject responses were found to be significantly different on factor 6 from graduate subjects. Comparing the two groups, it can be found that graduate subjects had a low score on that factor, while undergraduate subjects

TABLE 6. Mean concept ratings covered by factors 1, 2, 3, and 6 from different educational and professional status.

Factors and variables (ratings)	Group mean values: L*	G	U	H
Factor 1: Purposefulness of querying	**1.5****	**−0.21**	**−0.29**	**−0.17**
Information need: targeted/untargeted	3.5	1.7	1.7	1.8
Query: targeted/untargeted	3.7	1.7	1.8	2.2
Search: targeted/untargeted	4.7	1.9	1.6	1.5
Document content: targeted/untargeted	3.4	1.7	1.6	2.1
Factor 2: Applicability of data organization	**0.45**	**0.46**	**−0.21**	**−0.37**
Data structure: specific to IR systems/applicable to all Iss	4.0	4.0	3.1	3.1
Document content: specific to IR systems/applicable to all Iss	4.3	4.3	3.1	3.3
Feedback: specific to IR systems/applicable to all Iss	3.9	3.5	3.5	3.4
Interface: specific to IR systems/applicable to all Iss	3.7	4.1	3.5	3.2
Classification: specific to IR systems/applicable to all Iss	3.0	4.1	3.0	3.1
Factor 3: Function of querying	**−1.4**	**0.47**	**0.42**	**−0.13**
Information need: form/process	1.4	3.8	4.3	3.1
Query: form/process	1.8	4.0	3.8	3.3
Search: form/process	2.4	4.8	4.6	4.3
Factor 6: Function of data structure	**−0.33**	**−0.39**	**0.56**	**0.08**
Data structure: form/process	2.0	1.3	2.7	2.0
Interface: targeted/untargeted	1.5	2.1	2.5	2.4

* L: librarians G: graduate subjects U: undergraduate subjects H: high school subjects.
** Boldfaced numbers are average factor scores; normal numbers are average ratings.

TABLE 7. Results of ANOVA and Tukey test on factors for discipline.

Disciplines compared between:	Difference found on: $f_1^*(F = 12.01, df = 3, Pr = 0.0001^{**})$	$f_2(F = 3.11, df = 3, Pr = 0.033^{**})$	$f_3(F = 11.12, df = 3, Pr = 0.0001^{**})$	$f_7(F = 3.13, df = 3, Pr = 0.032^{***})$
Professional and sci./engi.	Yes***		Yes	
Professional and soc/hum.	Yes		Yes	
Professional and no-major	Yes		Yes	
Sci/Engi. and soc/hum.				Yes
Sci/Engi. and no-major		Yes		
Soc/Hum. and no-major				

*f_1 to f_7 stand for Factor 1 to Factor 7. Only the factors where a difference was found are listed. Refer to Table 3 for interpretations of these factors.
**Results of ANOVA test. Type III sum of squares was used.
***A difference, significant at $\alpha = 0.05$ level, was found between the two groups on the factor.

had a high score. Factor 6 included ratings of *data structure*: whether it is a *form* or a *process*. A high factor score indicated a preference over *process* and a low score meant a preference on form.

Effects of Academic Background on Factor Scores A significant difference among all groups (disciplines) was detected by the MANOVA test. The results of ANOVA on each factor for disciplines and Tukey test for comparisons between groups are exhibited in Table 7.

As showed in Table 7, significant differences were found on the first three factors and factor 7. Factors 1 and 3 reflected differences between professionals and student subjects. A difference between science/engineering subjects and "no major" subjects high school students was found on factor 2. Science/engineering majors were found significantly different from social science/humanities majors on factor 7.

It was also found that when divided by disciplines, only science/engineering subjects differed significantly from high school subjects. No significant difference was found between social science subjects and high school subjects.

Difference between Science/Engineering and Social Science/Humanities Subjects Average factor scores and ratings from different groups in terms of academic background are listed in Table 8.

Science/engineering and social science/humanities were found different on factor 7: purposefulness of browsing. When judging the activity of browsing, science/engineering subjects tended to browse purposefully, rating it as targeted. Social science/humanities subjects, on the other hand, considered browsing as a purposeless activity by rating it untargeted.

Differences between Science/Engineering Subjects and "No Major" Subjects "No major" subjects, as well as social science/humanities subjects, had an opposite score on

TABLE 8. Mean concept ratings covered by factors 1, 2, 3, and 7 from different disciplines.

Factors and variables (ratings)	Group mean values: P*	S	H	N
Factor 1: Purposefulness of querying	**1.52****	**−0.54**	**0.01**	**−0.17**
Information need: targeted/untargeted	3.5	1.3	2.1	1.8
Query: targeted/untargeted	3.7	1.5	1.9	2.2
Search: targeted/untargeted	4.7	1.5	2.0	1.5
Document content: targeted/untargeted	3.4	1.3	1.9	2.1
Factor 2: Applicability of data organization	**0.45**	**0.47**	**−0.1**	**−0.37**
Data structure: specific to IR systems/applicable to all Iss	4.0	3.9	3.1	3.1
Document content: specific to IR systems/applicable to all Iss	4.3	4.1	3.1	3.3
Feedback: specific to IR systems/applicable to all Iss	3.9	3.5	3.5	3.4
Interface: specific to IR systems/applicable to all Iss	3.7	4.3	3.5	3.2
Classification: specific to IR systems/applicable to all Iss	3.0	3.5	3.0	3.1
Factor 3: Function of querying	**−1.39**	**0.37**	**0.52**	**−0.13**
Information need: form/process	1.4	4.1	4.0	3.1
Query: form/process	1.8	3.6	4.2	3.3
Search: form/process	2.4	5.0	4.4	4.3
Factor 7: Purposefulness of browsing	**0.19**	**−0.63**	**0.34**	**0.09**
Browsing: targeted/untargeted	3.1	2.5	3.5	3.3

*P: professionals (librarians) S: sci./engineering subjects H: soc./humanities subjects N: no major (high school subjects).

TABLE 9. Results of ANOVA and Tukey test on factors for computer experience.

	Difference found on:		
Computer experience level compared:	$f_1^*(F = 3.76, df = 2, Pr = 0.029^{**})$	$f_2(F = 3.72, df = 2, Pr = 0.03^{**})$	$f_7(F = 3.66, df = 2, Pr = 0.032^{**})$
High and Medium			
High and Low	Yes***	Yes	
Medium and Low			Yes

* f_1, f_2, and f_7 stand for Factor 1, Factor 2, and Factor 7. Only the factors where a difference was found are listed. Refer to Table 3 for interpretations of these factors.

** Results of ANOVA test. Type III sum of squares was used.

*** A difference, significant at $\alpha = 0.05$ level, was found between the two groups on the factor.

factor 2 compared to that of science/engineering subjects. The former had a low average score while the latter had a high average score. A significant difference was detected between "no major" and science/engineering subjects. The difference reflected ratings on data organization concepts: whether they are applicable only to IR systems, or to all ISs. Science/engineering subjects seemed to consider the concepts as *applicable to all ISs* by holding a high score. "No major" subjects, however, tended to think of them as applicable only to IR systems, with a low score. The difference found here actually further revealed the source of the difference found between graduate and high school subjects on educational and professional status. Because there was no significant difference between social science/humanities and "no major" subjects, the difference between graduate and high school students was evidently the difference between science/engineering and high school subjects.

Differences between Professionals and Students Differences detected here between professionals and students were similar to those found between librarians and students on educational and professional status. Students contrasted with professionals on factors 1 and 3 in their factor scores, where professionals thought of querying concepts apparently as *untargeted* and *form,* subjects from other majors tended to consider these concepts as *targeted* and as *process*. These differences once again exhibited that librarians were a distinctive group, even from the perspective of academic background. It demonstrated how the professional training made differences in people's understanding of IR systems.

Effects of Computer Experience on Factor Scores The ANOVA test on each factor found that the differences occurred on factors 1, 2, and 7. Each level was compared against other levels separately using a Tukey test. The results of ANOVA and the Tukey test are presented in Table 9.

Significant differences were found between high and low level of subjects on factors 1 and 2. On factor 7, there was a significant difference between medium and low level. All the differences were significant at $p < 0.05$ *level. Subjects with low level of computer experience did differ significantly from those with high level, and from those with medium level of computer experience in their factor scores. No difference was found between high and medium level.*

Difference between High-Level and Low-Level Groups Corresponding groups' mean factor scores and corresponding concept ratings are listed in Table 10 to show the differences between groups.

The low-level group's factor scores sharply contrasted to the high-level group's, particularly on factors 1 and 2. The high-level group had high scores on both factors, which indicated that they considered querying concepts as *untargeted*, and data organization concepts as *applicable to all ISs*. The low-level group, on the other hand, had low average scores on both factors. They tended to think of query concepts as *targeted*, and data organization concepts as *specific to IR systems*.

Difference between Medium-Level and Low-Level Groups The low-level group was found different on factor 7 from the medium-level group. The two groups' scores showed that the medium group had a low score on the factor and the low group had a high score. Factor 7 was about ratings on the concept of *browsing:* whether it is *targeted* or *untargeted*. A low score meant in favor of *targeted,* while a high score pointed to *untargeted*. The difference here resembled the difference between science/engineering subjects and social science/humanities subjects. The former tended to browse purposefully, rating it as *targeted*; the latter, on the other hand, considered *browsing* as *untargeted*.

Discussion

The effects of three user characteristics on mental models are reflected in the differences among different types of users in their concept rating factor scores. The major findings of the study are the differences between librarians and students. Along the *educational and professional status*, librarians thought of querying concepts as *untargeted*, while all student subjects considered these concepts as *targeted*.

TABLE 10. Mean concept ratings covered by factors 1, 2 and 7 from different levels of computer experience.

Factors and variables (ratings)	Group mean values		
	H*	M	L
Factor 1: Purposefulness of querying	**0.38**	**−0.14**	**−0.41**
Information need: targeted/untargeted	2.1	1.9	1.6
Query: targeted/untargeted	2.5	2.0	1.9
Search: targeted/untargeted	2.7	1.6	1.4
Document content: targeted/untargeted	2.5	1.6	2.1
Factor 2: Applicability of data organization	**0.32**	**−0.00**	**−0.51**
Data structure: specific to IR systems/applicable to all ISs	3.9	3.2	3.1
Document content: specific to IR systems/applicable to all ISs	4.0	3.9	2.9
Feedback: specific to IR systems/applicable to all ISs	3.6	3.6	3.2
Interface: specific to IR systems/applicable to all ISs	3.8	3.7	3.1
Classification: specific to IR systems/applicable to all ISs	3.5	3.4	3.0
Factor 7: Purposefulness of browsing	**−0.01**	**−0.35**	**0.50**
Browsing: targeted/untargeted	3.2	2.7	4.1

*H: high level; M: medium level; L: low level.

Librarians related these concepts to *form*. Students, on the other hand, tended to perceive them as *process*.

The differences between librarians and students may be rooted in the fact that librarians were specially trained as information retrieval intermediaries, and they obtained skills needed for information retrieval through training and practice. In information retrieval, librarians normally function as intermediaries between end users and information retrieval systems. As professionals, they do search on behalf of end users, not for themselves. This characteristic, the separation of retrieving information and the desire to search, can probably explain why they tended to think the querying concepts *untargeted*. Because they search for other people and normally negotiate via forms with patrons about information needs, queries, and search results, librarians tended to judge these concepts as *form*, instead of *process*. Students, on the other hand, are end users. They usually search for themselves. When they search, they intend to find something to solve their own problems. That might be the reason that students thought querying concepts as *targeted* and as *process*.

Graduate students considered the concept *data structure* as *form*, differing from undergraduates' view of *process*. They also differed from high school students in their view of data organization concepts, including *data structure*, *document content*, *feedback*, and *classification*. They thought these concepts were *applicable to all information systems*. High school students, however, considered these concepts as *specific to information retrieval systems*. Graduate students' views were closer to librarians" than undergraduate and high school subjects were.

Taking the librarians' model as a relatively complete and appropriate model of the existing IR systems, the higher the education, the closer to the librarians' scores a student group's scores were. Graduate students' scores were closest to librarians'. Undergraduate students' scores were close to both graduate subjects' and high school students' scores. High school students' scores were most closely similar only to undergraduate students' scores. This finding is not surprising, because people who received more education would have more chances to learn and use IR systems.

Academic discipline played a significant role as well. The factor scores from science/engineering students contrasted with that from social science/humanities students on the concept of *browsing*. Science/engineering students considered *browsing* as *targeted*. Social science/humanities students, however, tended to think of it as *untargeted*. Science/engineering'es were similar to that of graduate students and social science/humanities's were similar to that of undergraduate students. One explanation to the difference might be that, science/engineering students are more practical and goal oriented. They do not browse unless they want to find some specific information. Social science/humanities subjects might like browsing even without something specific in mind or they might find out the needed information simply by browsing as the search process.

This *Untargeted* browsing attitude may be the reason why the users, particularly those with social sciences and humanities background tend to have an "evolving/berrypicking" like exploratory search model, as Bates (1989) proposed. According to this model, the user's search query "is an evolving one, rather than single and unchanging; and the search process is such that it follows a berrypicking pattern, instead of leading to a single best retrieval set." Bates (1989). It is not difficult to find a connection between "untargeted browsing" and the "berrypicking" pattern.

This finding may also be able to explain why users with science/engineering background tend to have better search performance than those with social sciences/humanities background (e.g., Kamala, 1991; Qiu, 1993). Conventional IR systems lack database browsing facilities. Therefore, social sciences/humanities users were in a disadvantageous position in using these systems than science/engineering users.

Computer Experience had a weak but significant effect as well. Subjects with different levels of computer experience differed in their factor scores. The high level group's scores were similar to that of librarians and the low-level group's were similar to that of undergraduate and high school students. Consequently, the low-level group tended to think of the querying concepts as *targeted.* The high level group had the opposite view. When judging data organization related concepts, the high level subjects thought of those concepts *applicable to all information systems* but the low-level subjects tended to think of them specific to information retrieval systems. The difference is similar to the difference between graduate and high school students. A difference was also found between the medium and low-level groups on concept *browsing*.

Although the above differences can be considered as reflections of that of educational and professional status, it may still be reasonable to conclude that a user's prior computer experience does have an impact on mental models, as found by some researchers (e.g., Borgman 1989). The findings suggest that users' computer experience should be taken into account when they are being trained for information retrieval. Similar concepts from other computer applications may be used for describing information retrieval tasks.

The reason of the differences here might be that the high-level group included all librarians and the low-level group consisted of mainly high school subjects (see Table 2). The high level group's scores were similar on the first two factors to librarians', and the low level's scores were the same as that of high school subjects' on the first two factors. The results indicated that librarians may be highly experienced computer users and high school students, on the contrary, can probably be novice users.

The difference between medium and low level of computer experience may again be caused by the fact that most members of the low-level group were high school subjects.

Language has been found to make a difference on user performance with a map interface (Chen et al., 1998): The native English speakers seemed to better understand the organization of the map and had an easier time using the map than nonnative speakers. However, first language had no significant impact on users' mental models in this study. Although IR tasks are heavily language dependent, the fact that the difference between the two groups was not significant enough in their factor scores contradicted the assumption of language's impact. The result might be due to the experimental conditions in this study, such as sample size and subject populations as well as their distributions in the sample. For example, most nonnative subjects recruited had been living in Canada for several years, and they had a reasonable comprehension of English language. Particularly, some of them moved to Canada as children and grew up in this English-speaking environment. It is hard to tell the difference between them and native English speakers in terms of language skills.

Implications

Implications for Information Science Research

Understanding the user diversity has been an important area of research in information science. The findings of the study enrich our knowledge about users and extend our understanding about differences among different types of users.

Previous individual difference studies found search performance or pattern varies among users and the variance can be predicted from certain user characteristics. But the question why users with different characteristics would have different performance has not been addressed by such studies. The findings of the present study demonstrate that different types of users have different mental models, which can explain the differences in search performance or behavior.

Research in IR interaction provides frameworks for examining users' roles in IR. However, little attention has been paid to individual difference issues. Differences in mental models may result in different interaction patterns. For example, the difference found in this study between social sciences/humanities and science/engineering subjects in their views on browsing (*targeted/untargeted*) suggests they would interact with IR systems differently, which has been evidenced by previous studies. Further research in IR interaction may need to consider the differences so that more complete interactive IR models can be developed. In addition, many complex aspects of interactive processes are still not fully understood (Spink & Saracevic, 1997). By exploring mental models, we may be able to understand users' interaction bahavior.

For mental model studies, this study suggests a more fruitful approach: associate models with known user characteristics so that differences can be easily identified. The power of mental models is to predict user performance from their models. By associating models with user characteristics that are established ways to categorize computer users, it becomes easier to assess differences and thus easier to predict differences in performance. This study found that three user characteristics had effects on mental models. Using the same approach the relationships between models and other user characteristics can be studied.

The methodology used in this study, particularly the repertory grid technique for obtaining and representing mental models, may be useful for future mental model research. The power of the technique stems from the following facts: (1) subjects do not have to use words to describe their mental models; (2) mental models can be represented in a unified format; and (3) mental models can be quantitatively analyzed. Consequently, quantitative comparison between different models becomes possible. The factors identified in this research may form a useful initial characterization of mental models. We hope by using the method that there will be more studies to validate and extend this characterization of attitudes towards information retrieval.

Implications for IR System Designs

Previous individual differences studies in IR demonstrate users can be distinguished on certain characteristics, yet such studies offer little predictive power for systems design that can best suit particular user groups (Dillon & Watson, 1996). One of the reasons is that such studies did not reveal why different characteristics would have different performance or behavior. Even if a user can be easily categorized into a group, if no mental models available, system features targeting at the group are hard to built. The general implication of the research reported here is that the differences in mental models in terms of user characteristics may form the basis for designing for heterogeneous user groups. Once the system identifies a user, the system can present the interface or feature(s), which are designed based on this type of users' views towards the system, to the user to accommodate to the user's specific anticipations and search plan.

Specifically, for example, the following findings may be useful for designs of IR systems. The first is that *query* was considered as "process" by student subjects rather than "form" by librarians. We did not ask subjects to elaborate what kind of "process" it would be. A process apparently involves more than one step, and it implicitly is an interactive process. This finding revealed that when end-users use IR systems they expect to conduct search as a process rather than a single step of entry of query. To serve end-users, IR systems should be designed to meet this anticipation rather than supporting professional librarians' "form" view, which normally requires submitting all necessary query-related information to the system at once.

The other example is that *browsing* was considered as "targeted" and "untargeted" by subjects with different academic backgrounds, respectively. It suggests that IR systems should be designed to combine direct search and browsing functions so that users with different academic backgrounds can both be accommodated. Conventional IR systems normally provide only direct search function, which may be in favor of the users with science/engineering background. To support social sciences/humanities background users, there is a need to add in an interface that will allow users to browse the databases. Yahoo! is a good example of such search engines: it facilitates browsing through its directories, and it also supports direct search by allowing the user to submit a search query. Among other reasons, this combination of search and browsing mode to support different types of users might explain why Yahoo is the number one search engine[1] on the Internet.

Conclusions and Future Research

To associate mental models with user characteristics, this study investigated the effects of four user characteristics on users' mental models of IR systems. The mental models were elicited by using the repertory grid technique, and were represented by concept ratings. Among the four characteristics investigated, educational and professional status, academic discipline, and computing experience had significant effects on users' mental models. The differences between user groups found in this study may explain individual differences in search performance/behavior found by some previous user studies. However, due to the novelty of the methodology used in this study and the exploratory nature of the study, the findings may need to be taken as suggestive.

There may be other user characteristics that are influential to users' mental models. Other user characteristics can be further studied so that interface designs could target at different levels and types of users based on their particular mental models (Sugar, 1995).

The study reported here concentrated only on mental models and how they differ between groups. Studies are needed that will directly connect models to both behaviour and search performance measures. This relationship may be difficult to establish, and may form the ultimate test of how useful mental model assessment and user differentiation techniques can be in informing the design of information seeking and exploration systems.

There are other methods for studying mental models, such as thinking aloud. RGT is only one of them. Each one has its own features. Further research using RGT on mental models is needed to test the robustness of the methodology used in this study. It will also be interesting to compare different methods for eliciting and representing mental models.

Acknowledments

We would like to express our special thanks and appreciation to Professor Charles Meadow, who inspired and motivated this research and provided considerable suggestions and the volunteers who participated in this study. Our thanks also go to the anonymous reviewers for their valuable comments on an earlier draft of this article.

References

Allen, B.L. (2000). Individual differences and conundrums of user-centered design: Two experiments. Journal of the American Society for Information Science, 51(6), 508–520.

Allen, B.L. (1996). Information tasks: Toward a user-centered approach to information systems. New York: Academic Press.

Bannister, D. (1968). The evaluation of personal constructs. London: Academic Press.

Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review, 13(5), 407–424.

Belkin, N.J., et al. (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., & Vickery, A. (1985). Interaction in information systems. London: The British Library.

Reuters reported on January 9, 2000, that Yahoo currently services 56 percent of all on-line search queries, with its nearest competitor, CMGI Inc.'s Alta Vista accounting for about 12%.

Bellardo, T. (1985). An investigation of online searcher traits and their relationship to search outcome. Journal of the American Society for Information Science, 36(4), 241–250.

Borgman, C.L. (1989). All users of information retrieval systems are not created equal: An exploration into individual Differences. Information Processing & Management, 25(3), 237–251.

Borgman, C.L. (1986). The user's mental model of an information retrieval system. International Journal of Man–Machine Studies, 24, 4764.

Borgman, C.L. (1984a). Psychological research in human–computer interaction. Annual Review of Information Science and Technology, 19, 33–64.

Borgman, C.L. (1984b). The user's mental model of an information retrieval: Effects on performance. Ph.D Thesis; DA 8408258, Stanford University, Stanford, CA.

Boyce, B.R., Meadow, C.T., & Kraft, D.H. (1994). Measurement in information science. San Diego: Academic Press.

Briggs, P. (1987). Usability assessment for the office: Methodological choices and their implications. In M. Frese et al. (Ed.), Psychological issues of human–computer interaction in the work place (pp. 381–402). Amsterdam: Elsevier Science Publishers.

Carroll, J.M., & Olson, J.R. (1988). Mental models in humancomputer interaction. In M. Helander (Ed.), Handbook of human–computer interaction (pp. 45–65). Amsterdam: Elsevier Science Publishers.

Charoenkitkarn, N. (1996). The effect of markup-querying on search pattern and performance in large-scale text retrieval. Unpublished Ph.D Dissertation, Department of Inductrial Engneering, University of Toronto.

Chen, H. et al. (1998). Internet browsing and searching: User evaluations of category map and concept space techniques. Journal of the American Society for Information Science, 49(7), 582–603.

Cool, C. et al. (1996). Information seeking behavior in new searching environments. In P. Ingwersen & N.O. Pors (Eds.), Proceedings of CoLIS 2, pp. 403–416.

Dillon, A., and Watson, C. (1996). User analysis in HCI—The historical lessons from individual differences research. International Journal of Human–Computer Studies, 45, 619–637.

Dimitroff, A. (1992). Mental models theory and search outcome in a bibliographic retrieval system. Library and Information Science Research, 14, 141–156.

Egan, D.E. (1988). Individual differences in human-computer interaction. In M.E. Helander (Ed.), Handbook of human–computer interaction (pp. 543–568). North Holland: Elsevier Science Publishers B.V.

Faulkner, C. (1998). The essence of human–computer interaction. London: Prentice Hall.

Fenichel, C.H. (1981). Online searching: measures that discriminate among users with different types of experiences. Journal of the American Society for Information Science, 32, 23–32.

Fransella, F., & Bannister, D. (1977). A manual for repertory grid technique. London: Academic Press.

Gaines, B.R., & Shaw, M.L.G. (1991). Basing knowledge acquisition tools in personal construct psychology. Research Report No. 91/453/37., Department of Computer Science, University of Calgary.

Halasz, F., & Moran, T.P. (1983). Mental models and problem solving in using a calculator. CHI'83 Proceedings pp. 21–26.

Ingwersen, P. (1992). Information retrieval interaction. London: Talyor Graham Publishing.

Kamala, T.N. (1991). Individual differences in the use of CD ROM databases. Unpublished Ph.D. dissertation, University of Hawaii at Manna, Honolulu, HI.

Kelly, G.A. (1955). The psychology of personal constructs. New York: Norton.

Kerr, S.T. (1990). Wayfinding in an electronic database: The relative importance of navigational cues vs. mental models. Information Processing & Management, 26(4), 511–523.

Kieras, D., & Polson, P.G. (1985). An approach to the formal analysis of user complexity. International Journal of Man–Machine Studies, 22, 365–394.

Latta, G.F., & Swigger, K. (1992). Validation of the repertory grid for use in modelling knowledge. Journal of the American Society for Information Science, 43(2), 115–129.

Marchionini, G., et al. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. Library and Information Science Research, 15, 35–69.

Norman, D.A. (1990). The design of everyday things. New York: Doubleday.

Norman, D.A. (1983). Some observations on mental models. In R.M. Baeck & W.A.S. Buxton (Eds.), Readings in human–computer interaction: A multidisciplinary approach (pp. 241–2441).

Preece, J., et al. (1994). Human–computer interaction. Wokingham, UK: Addison-Wesley Pub. Co.

Qiu, L. (1993). Markov models of search state patterns in a hypertext information retrieval system. Journal of the American Society for Information Science, 44(7), 413–427.

Robins, D. (2000). Interactive information retrieval: Context and basic notions. Informing Science, 3(2), 57–61.

Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. Proc of ASIS-1997 (pp. 313–327).

Savage-Knepshield, P.A., & Belkin, J.H. (1999). Interaction in information retrieval: Trends over time. Journal of the American Society for Information Science, 50(12), 1067–1082.

Sebrechts, M.M. et al. (1990). Integrating modeling: Changes in mental models during learning. In S.P. Robertson et al. (Eds.), Cognition, computing and coorperation (pp. 338–385). Norwood, NJ: Ablex Pub. Corp.

Shaw, M.L.G. (1980). On becoming a personal scientist. London: Academic Press.

Spink, A., & Saracevic, T. (1997). Interaction in information retrieval: Selection and effectiveness of search terms. Journal of the American Society for Information Science, 48(8), 741–761.

Sugar, W. (1995). User-centered perspective of information retrieval research and analysis methods. Annual Review of Information Science and Technology, 30, 77–109.

Voorhees, E.M., & Harman, D.K. (Eds.). (1998). Proceedings of the Sixth Text Retrieval Conference (TREC-6). (NIST Special Publication 500-240). Washington, DC: Government Printing Office.

Woelfl, N.N. (1984). Individual differences in online search behavior: The effect of learning styles and cognitive abilities on process and outcome. Dissertation. Cleveland, OH: Case Western University.

Yee, I.H. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 44(3), 161–174.

Zhang, X. (1998). A study of the effects of user characteristics on mental models of information retrieval systems. Unpublished Ph.D Dissertation, Faculty of Information Studies, University of Toronto.