©2008

YOO JIN HA

ALL RIGHTS RESERVED

ACCESSING AND USING MULTILANGUAGE INFORMATION BY USERS SEARCHING IN DIFFERENT INFORMATION RETRIEVAL

SYSTEMS

by

YOO JIN HA

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

Professor Daniel O. O'Connor

and approved by

;

:

samel 0.0 Corns

Carol C. Kakeran ynn Silipique Corneway Carol C. Kaktthan

New Brunswick, New Jersey

May 2008

ABSTRACT OF THE DISSERTATION

ACCESSING AND USING MULTILANGUAGE INFORMATION BY USERS SEARCHING IN DIFFERENT INFORMATION RETRIEVAL SYSTEMS

By YOO JIN HA

Dissertation Director:

Daniel O. O'Connor

There is an underlying assumption in the exchange of scholarly information that knowledge will be transferred across country borders, cultures, and languages. It is this sharing of scholarly information is considered an essential pre-requisite necessary for the advancement of knowledge. Nonetheless, in the current English dominant environment of information retrieval (IR) systems, there are numerous obstacles confronting users who seek to access and use non-English information.

The purposes of this study are: to explore the information behaviors of those seeking non-English information; to identify difficulties of individuals' experiences when accessing and using non-English information in current IR systems; to develop an explanatory model determining how person characteristics, experiential knowledge, and situation factors influence search behaviors and evaluations of bibliographic information.

ii

Two separate studies are conducted to explore the above issues: an online questionnaire of users of multilanguage information retrieval systems; and an experiment with individuals accessing information on different topics using different languages and systems. The participants in these studies include academic researchers and library personnel and are individuals who regularly interact with Chinese, Japanese, Korean and English records via IR systems.

The survey and experiment participants note the lack of non-English access via indexing terms, the lack of non-English records in major online databases which index journals, the lack of English translation of abstracts, and the lack of coherent and understandable access to non-Roman language materials. The users of non-English information expect to have a system with cross language information retrieval functions providing clear access to full text non-English information. Importantly, having understandable bibliographic records are essential when individuals make decisions on their expected use of non-English documents.

The experiment data analyses reveal there are different IR system search behaviors by subjects' with different language backgrounds, professions, language knowledge, topic knowledge and its target language, especially comparing English with non-English searches. An explanatory model for non-English searching model was built based on various statistical analyses of experiment data. The model depicts the importance of statistically significant relationships among person characteristics and experiential knowledge which explain search behaviors and intention to use retrieved information when individuals seek non-English/non-Roman alphabet information.

iii

ACKNOWLEDGMENTS

It is with sincere gratitude that I give genuine thanks to the researchers and librarians across the U.S who participated in my online survey and experiment. These individuals gave of their time and energy to make this research effort more meaningful. They approached the research with a seriousness of purpose and they went the extra mile to provide substantive comments on their experiences in using multilanguage information retrieval systems.

My advisor, Dr. Dan O'Connor, also deserves my lasting appreciation for the education he has provided to me from the first day I entered the Rutgers doctoral program. He has truly inspired me and it is my goal to emulate his fine example as a teacher, researcher, and active participant in our discipline. I am pleased to offer this work as an extension and continuation of the work begun by Professor O'Connor's own dissertation. It is with appropriate recognition when I state that my dissertation would not have been completed as it now stands without Professor O'Connor's continuous encouragement and support.

I also need to express resounding gratitude to my dissertation committee members, Professor Carol Kuhlthau, Professor Tefko Saracevic, and Dr. Lynn Connaway, a consulting research scientist of OCLC (Online Computer Library Center), for their critical comments to improve my dissertation. My contact with these individuals during my studies has helped me learn how define problems, think about ideas, and grapple with issues.

iv

I would like to acknowledge many people for helping me during my doctoral work. It is with a sincere heart that I recognize the encouragement and assistance given to me by Joan Chabrak who helps administer our doctoral program at Rutgers. I also thank SCILS faculty members who provided me with critical knowledge and who instilled in me a concern for and commitment to the users of our information systems. Our faculty broadened the scope of my horizon and gave my research interests a meaningful focus.

My fellow graduate students at Rutgers SCILS deserve my lasting appreciation as colleagues and friends who challenge each other intellectually and who share all the joys and fears of going through a doctoral program. Special recognition goes to Giyeong Kim and Eunju Oh who took care of me like my brother and sister since my arrival at Rutgers; they have guided and helped me to the successful completion of my degree. They listened to me and shared all my sorrow and happiness.

Finally I would like to offer a very special thanks to my family. I appreciate deeply how my parents provided me with unconditional faith and love during my long graduate work so far away from home. I also value the support given me by my sister, my two brothers and their respective families. My deep appreciation goes to my husband and his family for their steadfast encouragement. My husband, Jin-Hong Ahn, has traveled this journey with me and it is he who has been there to inspire, encourage, and support me to complete this work.

This dissertation is dedicated to my father, Jeong-Sun Ha, and mother, Gye-Soon Kim, who never stopped their love and belief in their little daughter.

V

Abstract	ii
Acknowledgement	iv
List of Tables	xi
List of Figures	xiv
Chapter 1: Introduction	1
1.1. Overview of this study	1
1.2. Current issues of bibliographic information	2
1.3. Global environment	6
1.4. Problem statement	8
1.5. Purposes of this study	
1.6. Overview of an experimental model	11
Chapter 2: Literature Review	15
2.1. Overview of literature reviews	15
2.2. Access to non-Roman alphabet information	
2.2.1. Global bibliographic information development	
2.2.2. CJK Romanization issues	
2.2.3. Availability of non-English information	
2.3. User information seeking behavior	
2.3.1. Information seeking behavior theories in general	
2.3.2. User information seeking behavior research in CLIR	
2.3.3. Non-English information users' information seeking behavior	
2.4. User's information search behavior	32

TABLE OF CONTENTS

2.4.1. Interactive IR models	32
2.4.2. User information search behavior in interactive IR systems	35
2.4.2.1. Studies in monolingual IR systems	35
2.4.2.2. Studies in CLIR systems	37
2.4.3. Language preference in query formulation	40
2.5. Bibliographic record evaluation and its selection by users	42
2.5.1. Relevance in general	42
2.5.2. Intention to use information based on bibliographic record	49
2.6. Summary of literature review	51
Chapter 3: Theoretical proposition and research questions	52
3.1. Theoretical background	52
3.2. Theoretical proposition	53
3.3. Research Questions	57
3.4. Premises/hypotheses for each research question	59
Chapter 4: Methodology of data collection	62
4.1. Overview	62
4.2. Experiment with observation and interview	63
4.2.1. Overview of experiment	63
4.2.2. Pre-test for experiment	65
4.2.3. Sample for experiment	67
4.2.4. Experimental design	68
4.2.5. Topics for experiment	69
4.2.6. Systems for experiment	72

4.2.7. Language assigned for this study's experiment	'3
4.2.8. Procedures used in the experiment7	'4
4.3. Online survey7	'6
4.3.1. Overview of online survey7	6'
4.3.2. Pre-test for online survey7	7
4.3.3. Sample and procedures for online questionnaire7	'8
4.3.4. Questionnaire details	30
4.3.5. Survey participants' background information8	\$1
4.4. Data analysis methods	3
Chapter 5: Data Analysis Results	37
5.1. Research question 1	37
5.2. Research question 2	25
5.3. Research question 3	;9
5.4. Research question 414	9
5.5. Research question 515	;3
5.6. Summary of data analysis result16	52
Chapter 6: Discussion, Implications, and Future Research16	<u>;</u> 9
6.1. An explanatory model of non-English information searching16	<u>;</u> 9
6.1.1. Effect of person's characteristics on searching for non-English	
Information17	'0
6.1.1.1. Different search behaviors by four language groups (CJKE)17	'0
6.1.1.2. Different search behaviors by different profession	'2
6.1.1.3. Bibliographic understand lever by profession and job experience17	'2

6.1.2. Effect of person's experiential knowledge on searching for
non-English information173
6.1.3. Effect of situation factors (Task, language, topic and system)
on searching for non-English-information173
6.1.3.1. Searching by different language173
6.1.3.2. Searching by three experiment systems
6.1.3.3. Different use behaviors by three different systems
6.1.4. Language choices and consideration in query formulation177
6.1.5. Bibliographic record understand level
6.1.5.1. Variables affect to bibliographic record understanding level
6.1.5.2. Reading different bibliographic records179
6.1.6. Non-English information users' relevance certainty judgments
(intention to use retrieved information)180
6.1.7. Query construction180
6.1.8. IR system preference
6.1.9. Satisfaction of the search
6.2. Current IR system use
6.2.1. System use when searching for non-Roman alphabet information
6.2.2. Users' needs and expectations when accessing non-English information185
6.2.3. Issues when accessing to non- English information and the effects
6.3. Validity and Reliability of this study with its limitation
6.4. Additional Implications for future research
6.5. Conclusion

Appendices	196
Appendix 1: Experiment	197
1.1. Consent Form	197
1.2. Instruction form for experiment	199
1.3. Example of task sheet	201
1.4. Post task questionnaires	202
1.5. Exit Interview questions	203
1.6. Exit questionnaires	204
1.7. Note for each task	207
Appendix 2: Survey questionnaires	208
2.1. Email advertisement for survey	208
2.2. Survey questionnaire	210
Bibliography	226
Curriculum Vita	239

LIST OF TABLES

Table 1. Top 15 Online Populations by Country 16
Table 2. Premises/hypotheses for each research question 59
Table 3. Subjects' sample distribution by language and field
Table 4. Overall Design of experiment
Table 5. Subjects' self-generated topics and queries in the experiment
Table 6. Languages used when searching for this experiment
Table 7. Summary of survey participants' background information81
Table 8. Data analysis plan by each research question & hypothesis
Table 9. Survey participants' experience of online sources 89
Table 10. Search focus for non-English materials 92
Table 11. System use for different formats searching for non-English information93
Table 12. Reasons given for respondent's most used systems when
seeking non-English information94
Table 13. Users' need for non-English information
Table 14. Non-English users' expectation for system
Table 15. First language choice from experiment100
Table 16. A prefer search language for non English information in English system100
Table 17. A prefer search language for non English information in non-English
System101
Table 18. Descriptive Statistics for non-English searching 103
Table 19. Descriptive Statistics of search behaviors by native language104
Table 20. MANOVA test for search behavior by language background105

Table 21. Tests of Between-Subjects Effects for search behavior by
language background107
Table 22. Post hoc tests for search behavior by each language group108
Table 23. Chi-Square test of Boolean term use between queries in
non-English searching109
Table 24. Chi-Square test of Advanced search option use by users' background111
Table 25. ANOVA test for search behaviors by research vs. library professionals112
Table 26. Descriptive for search behaviors by research vs. library professionals112
Table 27. Chi-Square Tests for advanced option use by researchers vs. library
professional113
Table 28. Correlation test of language knowledge with search behaviors 114
Table 29. Multivariate Tests(c) for search behaviors by experiential knowledge
Table 30. Tests of Between-Subjects Effects for search behaviors by experiential
knowledge117
Table 31. Descriptive statistics of search behaviors by English vs. non-English
Searches120
Table 32. ANOVA test for search behaviors by English vs. non-English searches121
Table 33. ANOVA test of search behaviors by different language searches
Table 34. ANOVA test of Satisfaction of the search result 124
Table 35. Issues in accessing non-English information
Table 36. Romanization issues 127
Table 37. Observation note of experiment subjects' system use
Table 38. Descriptives of search behaviors by three systems

Table 39. ANOVA tests for search behaviors by three systems	132
Table 40. Descriptives of Search time by three systems	135
Table 41. ANOVA test of Search time (minutes) by three systems	136
Table 42. Chi-Square Tests of give-up by three systems	.137
Table 43. Regression test for bibliographic record understand level	139
Table 44. Coefficients(a) of each variable for explaining bibliographic	
record understand level	140
Table 45. Mean of four bibliographic records	141
Table 46. ANOVA test for Record understanding level by English vs.	
non-English search	143
Table 47. ANOVA tests for Record understand level by language search	144
Table 48. Multivariate Tests of for bibliographic understanding level by	
researchers vs. library professionals	146
Table 49. Descriptive Statistics for bibliographic understanding level by	
researchers vs. library professionals	147
Table 50. A logistic regression test for intention to use of bibliographic record	150
Table 51. Descriptives of Relevance of intention to use by English vs. non-English	
Search	152
Table 52. ANOVA test for Relevance of intention to use by English vs. non-English	
search	152
Table 53. Regression model of satisfaction with the search	.154
Table 54. A regression model for a give-up search	156
Table 55. Summary of major data analyses results	162

LIST OF FIGURES

Figure 1. Experimental Model	12
Figure 2. A theoretical proposition	56
Figure 3. Boolean logic uses by language background	110
Figure 4. Box and whisker plot of satisfaction by English vs. non-English search	hes124
Figure 5. System efficiency by three systems	134
Figure 6. Satisfaction of the search by three systems	134
Figure 7. Termination of search for the three systems	137
Figure 8. record understand level by different language search	145
Figure 9. Record understand level by English vs. non-English search	145
Figure 10. Distributions of four bibliographic understanding levels for	
researcher vs. library professional	148
Figure 11. An explanatory morel of searching for non-English information	
(all arrows showing)	157
Figure 12. Non-English information searching: a distilled, brief overview of	
the derived explanatory model	161
Figure 13. An explanatory morel of non-English information	169

Chapter 1: Introduction

1.1 Overview of this study

By its very purpose, scholarly information is meant to be shared based on the premise that knowledge advances as individuals work toward the common goal of developing theories and models. This sharing of knowledge is assumed to transcend country borders, cultures, and languages. Nonetheless, in the current information retrieval (IR) environment, there exist impediments and obstacles confronting users who seek to access and use online databases when seeking information in different languages.

These problems are exacerbated when users seek information written in languages not known to them. Difficulties are encountered at various stages in the search process: constructing query formulations, interpreting the core bibliographic record, and evaluating the document itself. Although there is much research on translation, there is a surprising dearth of studies addressing cross-language information retrieval issues, especially those driven by theory which focuses on the real users of such systems.

The purposes of this study are: 1) To explore non-English information seekers' information needs and study their information seeking behavior; 2) To identify what kinds of difficulties individuals experience when accessing and using non-English information from current information retrieval systems; 3) To determine if analyses of searches and information from users supports the need for new features to improve multi-language access to index language and document surrogates and records for non-English information; 4) To develop explanatory models determining how user characteristics, language, topic, and task influence query formulation and relevance evaluations of

1

bibliographic information.

Two separate studies are conducted here to explore the above issues: (1) an online questionnaire of users of multilanguage information retrieval systems; and, (2) an experiment with individuals accessing information on different topics using different languages. The participants in these studies include those who regularly use IR systems: academic researchers and library personnel. These individuals were those who sought non-English information and regularly interacted with Chinese, Japanese, Korean and English records via non-English databases where they conducted research requiring non-English information.

1.2 Current issues of bibliographic information

In January 2008, the Library of Congress (LC) announced a working report called "On the Record¹: Report of the Working Group on the Future of Bibliographic Control." This report was commissioned by Deanna Marcum of the LC with other library professionals in order to redefine bibliographic control that can embrace a wide array of changes in materials, mediums, roles, and diverse user communities in the bibliographic world. In many ways, this report recognizes the current issues being faced in the bibliographic control profession. One of the examples used in this report is language issues: "As keyword searching becomes increasingly prevalent, non-textual works and works in languages other than English are at risk of becoming less accessible or even inaccessible (p. 20)." The report notes various mechanisms to address language issues and it suggests that the burden for providing access to such materials be shared with other

The report can be found at <u>http://www.loc.gov/bibliographic-future/news/</u>.

stakeholders:

"The Working Group identified a number of areas that might lend themselves to greater cooperative attention. First, there may be opportunities to work with the abstracting and indexing community, which is increasingly interested in the ability to identify more precisely the authors represented in its indexes. It may also be possible for LC to work with foreign national libraries that are engaged in similar activities (p. 19)." "Internationally shared authority files will enhance access to non-English language materials, including those in non-Roman alphabets and scripts, and will encourage international sharing of information and data (p. 21)."

There is already a long history of collaboration and cooperation among libraries and other information organizations. This cooperative work in providing access to information through bibliographic and other secondary routes is primarily based on the logical reconstruction of an indexing, abstracting, or cataloging system. However, such structures do not suggest what is really needed to improve and address current problems, especially in gaining access to non-Roman alphabet information. In other words, international bibliographic cooperation based on a continuation of the current systems denies the reality that such systems may not now be seen by users as efficient or effective in the retrieval of information across languages and alphabets. Instead, a more realistic approach might be to conduct actual user studies which can identify and define real user's needs and concerns when seeking multilanguage information.

Bibliographic information, defined here in its broadest sense as "information about information", can be used to create access pointers to more primary information. If the access pointers, such as index terms and name of publisher, are restricted by language, then those seeking information will be denied a key opportunity to locate source material which might satisfy their information query. This is especially perplexing if the original intent of producing (or "professing") the information was done to be shared and to

expand knowledge. Currently, translation can be used to access information in other languages and the avenues to do this are limited but expanding. Translation, however, only partly addresses the complexities of going from one alphabet to another. Transliteration, the isomorphic linking of one alphabetic sound symbol to a symbol in another alphabet, is similar to translation — both are attempts to provide bridges to users to get from one language to another.

Important here are the concerns of the individual who searches for information across languages. There is an underlying assumption in the exchange of scholarly information that knowledge will be transferred across country borders, cultures, and languages. It is this sharing of scholarly information that is considered an essential prerequisite necessary for the advancement of knowledge, especially in the sciences (Otlet, 1890s; Storer, 1970; Cole & Cole, 1972; Anderson 1974). For example, the sharing of worldwide progress in such areas as biological and medical advancement is assumed to be essential to the advancement of scientific scholarship. An extension of this concept of sharing knowledge can be seen as a motivating factor behind the current movements toward open access and open source information.

Nonetheless, for scholarly information to be shared there needs to be in place mechanisms to facilitate access to bibliographic records, conventions to explicate summaries of those records, and availability of translation services when needed. But this latter opportunity cannot occur if scholars are unable to find or understand the indexing system or the core bibliographic record because it is in a different language and, possibly, in a different alphabet.

Examining the assumptions which underlie access to and understanding of

bibliographic information is a complex task since this can be explored using many different methodologies. Of equal concern is the use of data in languages not known by the searcher. User focused research can be useful to inform the creation of new IR systems or to refining existing systems. Hence, the design of investigations in these areas might begin from different perspectives:

- examining the actual use that scholars make of information via citation studies;
- observing how individuals seek information with emphasis on gaining access to material in languages other than their own, especially languages in alphabets the individual may not know;
- asking individuals how they gain access to and interpret bibliographic data beyond their own knowledge of particular languages.

This research study focuses on user's experiences and issues when accessing non-Roman alphabet information using IR systems. Special attention is given to the evaluation of non-Roman alphabet bibliographic records in terms of users' access to this information and the users' understanding of the retrieved information. This study also explores users' query and use behaviors when accessing current IR systems. As noted earlier, a key focus of this study is to determine if there are fundamental differences when users conduct searches for English and non-English information. Finally, this study offers theoretical and empirical models to explain the importance of certain person variables, task variables, and search environment variables which influence users' decision making when seeking information across languages and alphabets using selected IR systems.

1.3. Global environment

Current information users are not limited by their location to access information. There has been an enormous increase in information written in different languages by users from different countries and cultures. As already noted, the current open source/open access movement is also designed to share information globally.

About 69.9 % of the world's online populations are non-English speakers and this represents roughly two-thirds of all Internet users (World Internet Usage, as of February, 2008), As Chowdhury (2003) affirms, multilingual information retrieval has now become a major challenge to gaining access to the prolific information on the web (p.72). This challenge can be partitioned into several major components and this overview provides a structure for understanding the definition of and resolution for the symbolic representation of language differences between the query and the document and the environment in which the search occurs.

The information world has not met all of its assumptions. It had been assumed that information access and use would be facilitated for information seekers because of the benefits of advanced technological improvements (e.g., rapid increases in Internet and WWW access, the development of spanning languages such as XML, advances in language processing, proliferation of online databases, etc). Information users, especially researchers and scholars, would appear to enjoy the convenience of an information environment where they can obtain research information by accessing bibliographic information using current IR systems, such as online public access catalogs (OPACs), online journal databases (DBs), and online web search engines. These developments might be expected to allow scholarly and public users to access information on a global level.

It is further assumed that the bibliographic surrogate stands as a gateway to more extensive data such as reports, books, and journal information. It is not the purpose of this study to challenge that particular assumption. But it is the purpose here to grapple with how users achieve or fail to achieve access to information written and indexed in other languages or alphabets. A particular component of this is that the access tool such as a DB may emphasize a particular language. In such cases, how might users seek information which spans languages or alphabets?

Assumptions also extend to scholars who create and use information. Technological developments of the Web, the effects of globalization, and the international emphasis in academic scholarship are three contributors to creating demand for ethnically diverse peoples to study, work, and collaborate together across borders and continents. Accordingly, today's information needs articulated by potential users of the Web and library databases expand the need for access to resources written in languages not known to the individual conducting the search. Global sharing requires effective and efficient systems which allow users to search document collections in multiple languages even when they have little or no linguistic competence in the target languages. It is important to note that regardless of this particular assumption of global participation, there does not yet exist an appropriate and unified information retrieval system which provides multilingual or cross language access. Translation engines currently serve a valuable but limited role. Yet, there is no easy way for scholars to harness scholarly information in a language and alphabet free environment.

1.4. Problem statement

Along with improvements in information retrieval systems, there have been many efforts to advance access to information via surrogate systems which include indexing, bibliographic description, and summaries of documents. There are, however, still difficulties accessing and using current IR systems including OPACs, online DBs, and search engines, especially when users want to seek information written in languages beyond their knowledge. Ha conducted three pilot studies in this area: (1) a focus group interview in 2004; (2) an online survey in 2005; and, an experiment in 2006. The results of these studies raised salient issues when users with different language backgrounds accessed multilanguage information. These studies mainly used OCLC's WorldCat system, which is the largest cooperative bibliographic system in the world. The findings from these studies include:

1. bibliographic records were difficult to read due to language barriers, especially for content written in Romanized words which are intended to substitute for non-Roman languages (or transliteration);

2. the lack of an English abstract in each record, serving as a Rosetta stone, would enhance understanding of the content of documents and books;

3. lack of standardization in bibliographic description across Roman and non-Roman alphabets where it was discovered that translation would facilitate understanding of the record;

4. Cross Language Information Retrieval (CLIR) functions were shown to be needed because users wanted to input their queries in their vernacular language even if the results appeared in another language. Three additional issues emerge in addition to those discovered in the pilot studies and these also challenge current use of IR systems to access non-English scholarly information. First, very few of the major online databases which index journals provide for non-English access (via indexing terms) or non-English records (Whitney, 1999). Second, even for those who provide access to non-English records, there is lack of coherent and understandable access to non-Roman language materials (Davis & Livny, 1994; Anderson 1974). Third, of the studies examined for this research, updating and extending the work of noted studies (Whitney, 1999; Davis & Livny, 1994; Liebergot, 1994; Anderson 1974) to other languages has not occurred and this leaves the full extent of the lack of access to non-English records unknown and unavailable.

Currently major database system providers including EBSCO Host appear to be providing a multilingual interface and advanced multilingual access functions to their collections. Yet, these functions build on complicated advanced search protocols which may not provide users with easy use of such systems; additionally, users may not know the extent of the DB's coverage of multilingual materials. In this study, users' actual experiment with EBSCO Host was conducted. The test results revealed that the EBSCO Host system does not accept multilingual queries yet it does access non-English collections; furthermore, the system was found to be difficult to use, especially for nonexperienced users. With the exception of the PubMed database, most non-English bibliographic records that users received from their searches did not have English abstracts and this created major problems in the user's understanding of the record's bibliographic information.

Google launched a language tool function in May 2007 which allows users to

translate their queries from one language to another so they can search for target languages. Or users can use it as a translation tool to translate the entire text of website. Users in this study noted the weakness of the quality of translation but they also affirmed that this is a valuable starting point for cross language information retrieval. Specific analyses reporting these findings are reported in Chapter Five. This is mentioned here to punctuate the point that a user centered IR system will need to grapple with a multitude of issues as it attempts to provide meaningful access to information in different languages.

Certain CLIR and IR researches have addressed language issues and have proposed new systems to achieve some of these goals (Peters, 2000 & 2006). Yet, many studies did not investigate user concerns nor did they assess existing systems which were attempting to encompass multiple languages in order to provide pertinent access points (cross language or multilingual access) to bibliographic records. To achieve global access to information, system designers will need to address such requisites as indexing access and bibliographic description using multiple languages. They may also need to account for document surrogates in languages likely to be understood by the largest number of potential users.

1.5. Purposes of this study

- To explore non-English information seekers' information needs and study their information seeking behavior;
- To identify what kinds of difficulties individuals experience when accessing and using non-English information from current information retrieval systems;

- To determine if analyses of searches and information from users support the need for new features to improve multi-language access to index language and document surrogates and records for non-English information;
- To develop explanatory models determining how user characteristics, language, topic, and task influence query formulation and relevance evaluations of bibliographic information.

1.6. Overview of an experimental model

The experimental model proposed for this study depicts how a search process is partitioned where certain concepts are linked to explain information seeking behaviors when using multiple languages (see Figure 1. Experimental Model). Three major concepts include the influences of particular person, task, and other situational characteristics on query formation, database selection, and intended information use. Included in these considerations are: 1) users' information search behaviors; 2) access to non-Roman alphabet information; and 3) bibliographic record evaluation (relevance judgments).

The experimental model was derived from a proposition that "access and use of non-English language information is a function of:

1) a person's user characteristics (native language, job, age, etc);

 a person's experiential knowledge (language knowledge, system experiences, topic knowledge);

3) information opportunity (system components including language coverage and its retrieval functions);

4) bibliographic access mechanisms (language access for index terms, and users query construction protocols);

5) bibliographic record structure (testing if English could be used as a Rosetta stone linking languages through title and abstracts serving as surrogates for documents);

6) user's satisfaction with the system's retrieved search results; and,

7) user's relevance judgments as an intention to use the retrieved bibliographic record to obtain a document.

Figure 1. Experimental Model



Hypothesized relationships between and among variables include:

1. Non-Roman alphabet information users' information seeking behaviors

Person's characteristics and experiential knowledge (as independent variables (IV1) influence or explain variability in user's selections of information sources as a dependent variable (DV1), information search query formulation (DV2), bibliographic record evaluation (DV3), and selection of bibliographic record (DV4). Person's characteristics include age, education, native language, job experience. Person's experiential knowledge includes person's knowledge based on their experience of language, subject, system, perception of the system, and searching. [Note: Dependent Variables in Stage 1 become Independent Variables in Stage 2 of the model.]

2. Access to non-Roman alphabet bibliographic information

2.1. DB system policy (IV2) influences users' database selection (DV1) and bibliographic record selection (DV4). DB system policy includes language coverage, retrieval function, subject coverage, perception of database prestige, interface mechanisms, perceived ease of access, and availability of language services.

2.2. Information system search (query formulation) (IV3) influences

bibliographic information evaluation (DV3). Construction of query formulation, especially when searching for non-Roman alphabet information, is posited to influence the selection and evaluation of bibliographic records. Independent variables at this stage in the search process include language choice of the query, query length, number of query changes, search time, ease of query formulation, or support of translations across queries and their concomitant influences on records searched and retrieved.

3. Evaluation of bibliographic records

Person's characteristics and experiential knowledge (IV1) effect bibliographic information evaluation (DV3) and bibliographic record selection (DV4) as these influence intentions to actually use the retrieved information (DV5). The bibliographic information evaluation (IV4) influences record selection (DV4) for information use (DV5). Evaluation takes into consideration whether the record includes an English abstract, if the record encompasses the vernacular (or translated) language and English descriptors, and record and abstract description, whether vernacular or English.

The following literature review section is structured by the theoretical and empirical models. Research questions and hypotheses will be derived, respectively, from these models and from the literature review.

Chapter 2: Literature Review

2.1. Overview of literature reviews

This literature review section examines four different topical areas that are linked to each other under a common concept, the "users' point of view." This approach centers on the user of an information retrieval system who is seeking information across languages and alphabets. In some cases, the user's native language is a non-Roman alphabet language. It is estimated that the deep Web (or invisible or hidden web) is "500 times larger than the surface Web (He et al., 2007)." Surprisingly, almost one-third of the deep Web is indexed by Google and Yahoo (and with overlap accounted for, 37% of the deep Web is accessible by these two search engines) (He et al., 2007, p. 99). Important to the size of the Web is its complexity and the extent of languages used. Currently, English is a dominant language of the Web, but that is changing. Notice the distribution of Web users reported by comScore in the following table 1.

It can be assumed that English may not continue as the Web's dominant language as other countries provide more access to the Internet and to the Web itself. From Table 1, it can be seen that China, Japan, and South Korea have, together, the same number of defined Internet users as the United States. Hence, the proliferation of languages and access to information in those languages becomes a significant concern. This study has as its core premise that user preferences for access to multilanguage information may help shape the information retrieval system of the future. Google became a central access point to web information based on users' behaviors and preferences. A similar approach may be expected with multilanguage access. Hence, this study attempts to view such access from users' viewpoints by soliciting user feedback via a survey and by observing users and recording their patterns in an experimental setting. The study focuses on information retrieved in English, Chinese, Japanese, and Korean languages (the latter three languages are referred to as "CJK"). Other languages are also included as they appear in the citations retrieved from searches.

Worldwide – All Locations	Unique Visitors (000) [in millions]
Worldwide Total	694,260
United States	152,046
China	74,727
Japan	52,100
Germany	31,813
United Kingdom	30,190
South Korea	24,645
France	23,884
Canada	18,996
Italy	16,834
India	16,713
Brazil	13,186
Spain	12,452
Netherlands	10,969
Russia	10,833
Australia	9,735

Table 1. Top 15 Online Populations by Country

(Source:² comScore World Metrix. Among visitors age 15+ in March 2006)

The four concepts explored in this study which include the users' point of view are: 1) Access to non-Roman information; 2) Non-Roman alphabet users' information seeking behaviors; 3) Users' search behaviors in interactive IR systems; and 4) Evaluation of bibliographic information.

²) comScore announces "New worldwide online Universe estimate based on the world's largest, most representative sample and most robust methodology" (May 4, 2006). Retrieved March 3, 2008 from http://www.comscore.com/press/release.asp?press=849

The first part of this literature review covers access to non-Roman alphabet information. Global bibliographic information development and Romanization issues in CJK languages are contrasted with issues of availability of non-English information. The second section reviews user information seeking behavior literatures starting with an overview of information seeking behavior theories in general, and then moving to user studies in Cross Language Information Retrieval (CLIR). Then, non-English information users' information seeking behaviors will be examined. The third section, research about users' search behavior will start with an introductory review of current interactive IR models. Then, this review will examine users' search behavior studies in IR systems and language preference in users' query. Finally, relevance and intention to use information based on bibliographic record evaluation will be reviewed. An overview of the major topics in this literature review is:

2.2. Access to non-Roman alphabet information

- 2.2.1 Global bibliographic information development
- 2.2.2 CJK Romanization issues
- 2.2.3 Availability of non-English information
- 2.3. User information seeking behavior
 - 2.3.1. Information seeking behavior theories in general
 - 2.3.2. User information seeking behavior research in CLIR
 - 2.3.3. Non-English information users' information seeking behavior

2.4. User's information search behavior

- 2.4.1. Interactive IR models
- 2.4.2. User information search behavior in interactive IR systems
- 2.4.3. Language preference in query formulation

2.5. Bibliographic record evaluation and its selection by users

- 2.5.1. Relevance in general
- 2.5.2. Intention to use information based on bibliographic record

2.2. Access to non-Roman alphabet information

2.2.1. Global bibliographic information development

Paul Otlet and his colleague Henri La Fontaine were the first creators of an international organization working for global documentation in early 1890s. They were committed to make a union bibliography and universal classification system which could be used on a worldwide level (Rayward, 1975). They were true proponents of today's globalization. Otlet tried to make a synthesized and systematic list of science works. His goal was to create a separate entry from the work itself where each item of the list could provide an access point to researchers in a worldwide catalog. From this idea, today's bibliographic tools, such as the catalog, index and abstract systems could be created (Rayward, 1975, p. 28-32).

This creative thinking of Otlet and LaFontaine made it possible to design the principles underlying today's union catalogs including that of the Online Computer Library Center (OCLC) /WorldCat. It is telling that library and information professionals have been trying to follow Paul Otlet's spirit, embodying universal bibliographic description to create accessible and understandable information about other information. It is also a very important claim not only to make a list of access points but also to propose standards for a bibliographic record which could be used worldwide.

The advent of Machine Readable Cataloging (MARC) in 1967 accelerated the sharing of bibliographic information on an international basis. In 1969, "the International Meeting of Cataloging Experts" was held in Copenhagen, and at this meeting, an international working group was established to develop a standard order and content for monographic description (Chan, p. 42, 1994). As its first product, the workgroup issued a document titled ISBD (M): *International Standard Bibliographic Description (for Single Volume and Multi-Volume Monographic Publications)* in 1971. This document was officially published in 1974, and it received much notice from the international community. For example, clarification of ambiguities and corrections were recommended by the International Federation of Library Associations and Institution (IFLA) after its conference in 1973.

The creation of *the International Standard Bibliographic Description for Monograph Publications* provided for compatible descriptive cataloguing worldwide that facilitated the international exchange of bibliographic records between national bibliographic agencies and throughout the international library and information community. This monographic standard soon led to the development of similar International standards for other types of materials and formats. IFLA, as an international professional body for the world library community, has been producing various guidelines and standards for regularizing library work while increasing bibliographic control over acquired materials, such as establishing an International Office for *Universal Bibliographic Control (UBC)* in 1974 (Anderson, 1974, pp. 11) and creating its *Functional Requirements for Bibliographic Records (FRBR)* in 1990 (IFLA, 1990). So the history of standardization in the library world has been a core concern for over 40 years and many cooperative movements have continued to pursue a long term goal of global information sharing.

Recently, current concerns for multilingual access to bibliographic information in catalogs were recognized by The Library of Congress. On December 2006, the Library Congress established a working group on the future of bibliographic control to address current issues of library work. The working group members are compiled of professionals in Library and Information Science (LIS) field, Microsoft, Google, and OCLC. Deanna Marcum and the working group recently released "On the Record: Report of the Working Group on the Future of Bibliographic Control (January 9, 2008)." ³ The working group urges a redefinition of bibliographic control that embraces a wide array of materials, diverse user communities, myriad venues where information is sought, and many sources of metadata for organizing collections for discovery and use. The report mentions the importance of the language issue: "As keyword searching becomes increasingly prevalent, non-textual works and works in languages other than English are at risk of becoming less accessible or even inaccessible" (p. 20). Google, EBSCO, Elsevier, and others are just now considering and/or enhancing various translation services for queries and for results of searches.

2.2.2. CJK Romanization issues

Since the early 1980s, when bibliographic records were entered with original vernacular data by the two major bibliographic utilities, OCLC and the Research Libraries Information Network (RLIN), non-Roman scripts in OPACs were used according to an agreement on the process/procedures for transliteration, where symbols would be transliterated to alphabetic characters and vice-versa (Taylor, 2000, p. 462-472; Shaker, 2002, p.3). In 1987, a meeting that discussed non-Roman alphabet problems was held by IFLA in Tokyo. The results of that meeting were summarized and published in a work titled *Automated Systems for Access to Multilingual and Multiscript Library*

³ The report can be found at <u>http://www.loc.gov/bibliographic-future/news/</u>.

Materials: Problems and Solutions (Bossmeyer & Massil, 1987). The main topics discussed were the need for standardization with all types of scripts. Of particular concern were ideographical scripts and the need for technical systems to support vernacular data.

IFLA continues to have its meetings address these concerns to cover multilanguage and multi-scripts practices in the provision of catalog information. In 1993, the three IFLA Sections held a joint meeting to combine these separate groups: Information Technology, Library Services to Multicultural Populations, and the Division on Bibliographic Control. The meeting's main theme was to focus on multilingual and multi-script problems in organizing and providing access to catalog information. Unicode issues were discussed in a 1995 meeting to solve the standardization problems in different character sets (IFLA, 1993, 1995).

Research has continued to explore the role of Romanization in cataloging and the increased use of vernacular records. Studies in this area have focused on the development of logical principles with concomitant attention to cataloging rules and standardization guidelines. Among non-Roman scripts issues, there has been active research done on Romanization by LC and OCLC: in Chinese, Japanese, and Korean (CJK) scripts area: (Zeng, 1992 & 1992; Arsenault, 2005; Zhang & Zeng, 1998; Shin 2003).

There were fewer studies conducted that considered standardization issues as they related to specific language areas where international scholars wanted a uniform system to describe a published work. Zhang and Zeng (1998) examined practical problems using the Unicode[™] Standard in library applications to examine standardization in bibliographic description, specifically in CJK information processing practices. Zeng
(1991) conducted research comparing OCLC CJK system with the RLIN CJK system. The conclusion of this study focused on the CJK thesaurus used in the creation of records and it emphasized the need for strict adherence to standards.

Another evaluation of the OCLC CJK Plus system was done by Jeong (1998). He conducted an experiment with 32 participants from Chinese, Taiwanese, Japanese and Korean language backgrounds. Jeong tried to focus on end users' searches using three different versions of the OCLC CJK Plus' search mechanism (Roman-derived search, Roman title-phrase search and vernacular search). Note that these were all cataloger specific systems not available to end-users. Even so, the transliteration issues of the catalog users were not a focal point of this research. The experiment did not allow system users to access the database using their preferred language

Park (2001) also addressed the Romanization issue with a special attention to using the "McCune-Reischauer (MR)" for the Korean language in the current bibliographic utilities, such as the OCLC CJK system. She claimed there are many problems in using the "McCune-Reischauer (MR)" for the Korean language in current bibliographic utilities, such as the OCLC CJK system. The MR is a Romanization scheme for Korean and it is still used. Park identified the difficulty in creating a system with less ambiguity using several real Korean bibliographic utilities made by MR. There have been attempts to make new software available, but it has not been released yet. This, too, may lead to another standardization issue.

Shaker's dissertation (2002) investigated how current academic library systems can support non-Roman materials and what should be considered in order to make it possible to have vernacular characters in those systems. That work covers various transliteration issues related to current cataloging practice as well as examining many different languages used in bibliographies (i.e., Cyrillic, Arabic, Hebrew, and CJK). Ha (2006) examined problems accessing and using Multilanguage materials from the endusers' perspective. Users indicated that transliterated (Romanized) information could not be understood and that gaining access to records was inconsistent. This could indicate that the good intentions of those who created mechanisms to facilitate access had, in effect, created systems which increased user confusion and frustration. Clearly, this also pointed to the need to conduct additional research to find out what users were experiencing when their searches involved transliterated information.

2.2.3. Availability of non-English information

Language support in library collection or database coverage plays a very important role in identifying and providing access points to information. Allan (1990) examined four different databases (Bibliofile, Dialog, OCLC, and RLIN) using titles search protocols (over 1,000 items) in order to determine whether their performance can support increases in users' demands at that time (p.393). She conducted the search using both English and foreign languages to see whether the databases can support the foreign language. Although the foreign language portion was very small (about 5%), it represents the initial concerns of an important access area.

Language distributions in major online bibliographic databases, particularly in language coverage, were explored during 1970-1984 by Whitney (1990). Eight databases (BIOSIS, Chemical Abstracts, GeoRef, MEDLINE, Criminal Justice, Oceanic Abstracts, PAIS, PsyInfo) on DIALOG were chosen. Whitney reported there was obvious evidence to drop non-English materials in eight major databases during the 1970s to 80s. Whitney claimed that English was too dominant in the distribution of database languages and that this actually worked to inhibit access to non-English materials.

A related and important study was done by Davis and Livny (1994) who identified the difficulty in accessing Japanese science information by users who do not knowing Japanese. They examined citations in the Japan Information Center of Science and Technology which is an English version of that database (JICST-E File database). They found that around 90% of the number of citations is identified by Japanese but not available to an English language query. That is, the JICST database does not have a cross language information retrieval function or any translation help so that much information was only available for those who knew Japanese. Liebergot (1994) did his dissertation work by interviewing researchers of physics laboratories in Japan. He also found the availability of a database with appropriate information and with language support is a significant factor for science users.

Vilar and Zumer (2005) evaluated the interface function of four different online databases for journals (ScienceDirect, EBSCO Host, ProQuest, Emerald) in 2004. At that time, only ProQuest provided a multilingual interface option and the authors claimed that it might negatively affect users who needed the language function. Currently (as of February 2008), among those four databases, there is still no database providing complete cross language retrieval functions and vernacular language search functions. Science Direct and ProQuest do allow for searches using romanized letters but not original scripts.

Not a long ago, Kellsey (2003) noted that current research in academic libraries

needed additional foreign language expertise in order to provide better access to meet increasing demands from users. Kellsey pointed out cooperative works (including interlibrary loan) have not gone far enough to support these needs and it was suggested that this problem be address in LIS education. This reinforces the need for providing effective access to international, multilanguage information during this global age.

Most recently, Kim (2006) tried to understand what kinds of users' perceptions affect users' acceptance of web-based subscription databases based on TAM (technology acceptance model) which is widely used to examine web user behavior or evaluation in the context of the World Wide Web. TAM theory comprised two belief constructs: "perceived usefulness" and "perceived ease of use." Kim tested a total of seven variables to determine if the variables influence the two beliefs, and then examined this to see if it finally affected the users' intention to use the database. The seven variables Kim tested are: subjective norm, job relevance, output quality, result demonstrability, user training, accessibility, and terminology clarity. Kim used a web survey method with undergraduate students asking them to compare three databases: the Education Resources Information Center (ERIC), PsyINFO, and Library Literature & Information Science. The result indicated that the 'terminology clarity' and 'accessibility' were the most important determinants for 'ease of use.' The effect of 'usefulness' belief in intended use was greater than the effect of 'ease of use' belief. Although this study focused on users' perception of using web-based databases, the study provided an overview of users' evaluation criteria about databases. This result confirmed that the importance of appropriate 'language' coverage both in the interface and the search can enhance users' database use experience.

2.3. User information seeking behavior

2.3.1. Information seeking behavior theories in general

Wilson (1999) defines information seeking behavior as "those activities a person may engage in when identifying his or her own needs for information, searching for such information in any way, and using or transferring that information" (p. 249). There have been extensive users' studies under the name of 'human information seeking behavior,' which have developed various models in order to assess a pattern of information behavior in certain environments, such as the "Anomalous States of Knowledge (ASK)" model which indicates users do not know exactly how to construct a query in an unknown area so the system needs to provide them with assistance (Belkin, 1980; Belkin & Oddy, 1982a, 1982b). An important depiction of the sequential decision making experience by users is addressed in Kuhlthau's "Information Search Process (ISP)" (Kuhlthau, 1983, 1988a, 1988b, 1993a) and Ellis's information seeking pattern model (Ellis, 1989, 1993). Work in related information seeking research include: "Sense-making" approach (Dervin, 1983, 1992); "berry-picking" (Bates, 1989); "Information Use Environment (IUE)" model (Taylor, 1991); and, user-oriented information retrieval research focusing on human behaviors (social and cognitive situations) as users interact with IR systems (Kuhlthau, 1993a; Kuhlthau, Spink, & Cool, 1992; Belkin & Vakkari, 1985; Vakkari, 2003).

Something as simple as a person using information becomes a complex layering of processes and changes as researchers explore the underlying structure in information seeking. An individual is assumed to have an information need which creates an internal motivation to behave in an information environment. Dervin (1983, 1992) addressed this

need as a process of "sense-making" characterized by an individual needing to resolve a problem. Kuhlthau (1993a) saw this instead as a process people engaged in when assigned tasks requiring information. However, Belkin (1980) began his investigations of this process after the need had been expressed and saw it as an interaction between query and information source with a continuous modification of the need as new information informed the person and acted to modify the query. Meanwhile, Pettigrew et al. (2001) attempted to define the sequence of the seeking process while Bates (1989) drew upon analogies to explain this process. Saracevic depicted the complexity of IR interaction in a stratified model where user and system components required a continuum of adaptation (Saracevic, 1997). Saracevic's model is also discussed later in this chapter. Note that these studies give important insights into user behavior and they offer promising variables to consider when representing how user interact with IR system in a multilanguage environment.

2.3.2. User information seeking behavior research in CLIR

Language related user studies can be found in Cross Language Information Retrieval (CLIR) research. CLIR is a retrieval system that operates with queries in one language to retrieve documents in other languages. It allows users to access information written in the user's languages of choice. CLIR has been studied from diverse research fields such as Information Retrieval (IR), Natural Language Processing (NLP), Machine Translation (MT), Linguistics, and Human-computer interaction (Oard et al., 1999).

Until recently, CLIR research has focused on physical system development, for example, development of translation techniques for query translation or content

(document) translation using various methods: Ontology, Machine translation lexicons, Bilingual dictionary and Corpora machine translation (Oard, 1997). Most of these studies have been done without considering actual users' use of the systems. Oard and Resnik (1999) argued that three disciplines — information retrieval, library science, and machine translation — should integrate their research in order to provide for a well-designed CLIR system (p. 364). A critical issue for the CLIR research is how to define and study users of bibliographic or IR systems. Saracevic (1999) gives emphasis to user aspects by stating that "real progress in information science and by extension in Information Retrieval will come when we put the 'human' in the process of building a system."

There have been few studies by IR user researchers about access to information spanning different languages of users, different languages of target documents to be retrieved, and access to language information unknown to users. The user studies in CLIR are still in a beginning stage to examine particular CLIR systems in terms of users' needs, preferences, and behaviors (Ogden et al., 1999; Petrelli et al., 2004). Ogden and his colleagues conducted one empirical user study asking who the users are and what their goals are. They postulated the potential ways a CLIR system may be used: 1) bilingual users who have good reading skills in their second language, but may have poor language productive skills. They, then, cannot express their information needs in their second language as well as they can in their first language. 2) monolingual users who have an interest in finding information written in other languages, but at the expense of much time and cost.

It is recognized that users have diverse backgrounds, and it is implied they may have different information seeking behaviors due to culture, language, or learned styles. Resnik & Oard (1999) discovered that due to the diverse backgrounds of CLIR users, a helpful interface is critical which can enhance formulating queries which, in turn, can interact with the system. They also claimed designing a sensible and simple display of retrieval results (surrogates) will be required based on users' studies

Petrelli et al. (2004) used a case study by observing and interviewing real users (10 subjects: business analysts, journalists, librarians, translators) at their work place. From these interviews, they identified a number of user requirements for CLIR systems. They found that search behaviors depended not only on user goals or purposes, but also on the language knowledge of the individual and the cognitive demands of the cross language task itself. Users indicated that they want to: search multiple languages simultaneously, to change query languages within the same search session, to support multilingual queries, to search by the most appropriate language they know for the task, which is not always their native language, and to filter results by language, genre, date, or other features (p. 928). They argued that designers of CLIR systems should examine cross-lingual information search tasks in real environments with real users to overcome the mismatch between user goals and system mechanisms.

There were system evaluation studies where the system was evaluated by real users (Chung et al., 2004; Qin et al., 2006). Chung et al. (2004) evaluated a portal CLIR system by studying real users. The portal system named "CBizPort: Chinese Business Intelligence Portal" is a meta-search engine for business information of China, Taiwan, and Hong Kong and they asked users to compare this system with other search engines from these three regions. They started to investigate problems in existing search engines which might not serve many non-English speaking Internet users. From the beginning of the experiment and the evaluation, they tried to focus on real users' needs and difficulties. Another study by Qin et al. (2006) asked domain experts to evaluate a multilingual web portal. They tested in both English and Chinese. Although these studies include users in their study, it did not attend to a detailed study of users' behaviors.

Most recently, Petrelli (2008) emphasized again the users' role as important in current interactive system evaluation. He and his colleagues worked on the Clarity Project and they tried to focus on user centered evaluation. In all, there have not been many studies involving how individuals seek and search multi language information databases, especially those where the query and retrieved documents are in languages involving different alphabets. Access to bibliographic information is known to have some connection to the users' culture and language. However, the degree to which these variables are important is not yet known. Many of CLIR studies still have focused on language translation in engineering aspects of creating CLIR systems.

2.3.3. Non-English information users' information seeking behavior

End users' online search behaviors were the object of a number of investigations shortly after the introduction of online catalogs in public and academic libraries in the late 1970s (Borgman, 1986, 1996; Borgman & Siegfried, 1992; Chen & Dhar, 1990; Taylor, 1984; Markey, 1983). Borgman (1996) pointed out although there have been improvements in the interface design of OPACs, end users are still encountering difficulties when they use these systems. Borgman claimed that it is because there is a lack of attention to user studies by system designers; furthermore, those who create systems should recognize the plight of the end user when seeking information. Since the 1990s, there have been many user studies examining information seeking behaviors and these studies have explored various users groups, by occupation, by social role, or by demographic groupings (detailed information behavior studies were well described in Donald O. Case's book (2002).

There has been little attention paid to exploring the information seeking behaviors of specific users who are not native English speakers. Little research has done with immigrants who use a specific library for their information seeking in the U.S. (Fisher, et al. 2003; Rho, 2002; Ganss, 1999; Berger, 1999; Gonzalez, 1999; Jones, 1999; Su & Conaway, 1995). Fisher et al. (2003) conducted case studies with non-English native users of New York Public Library. This study examined the users' information seeking behaviors, such as where they go to find specific information and how they use the library. Rho (2002) looked at Korean immigrants who live in a specific region in the U.S. and explored this with questions similar to those used by Fisher et al. It was found that language barriers and cultural differences inhibit individuals' access to and use of catalogs and information centers/libraries.

There have been few studies done of international graduate students who study in the United States (Liao, Mary, & Lio, 2007; Jeong, 2004). Liao et al. (2007) conducted an online survey to do a comparative study on information needs and information-seeking behaviors of international graduate students and American graduate students. Jeong (2004) had in-depth interviews with Korean graduate students in terms of information seeking in their everyday life. Again, this and similar studies indicate that the language barrier is a substantial factor affecting individuals' information seeking and use. These studies looked at specific groups' overall information seeking behaviors in their daily lives rather than focusing on their IR system access and use.

2.4. User's information search behavior

2.4.1. Interactive IR models

The concept of interaction in IR models has received much attention in recent research (Saracevic, 2007a, 2007b, 1997). The concept underlying interaction has to do with the intersection of a person with an information system. In earlier research, interaction was also studied as a statistical entity either by intention or by default. This notion of interaction grew from an awareness that the simple matching of a user's query with retrieved results failed to recognize the intersection of the human with the system which could account for important variability in explaining the results obtained. This recognizes that a fundamental element in IR is the interaction of the human and system with many researchers asserting that the users' query formulation and the users' interaction with the system becomes a defining characteristic of the information search process (Belkin et al., 1980).

There have been many interaction models (e.g. O'Connor, 1978; Ingwersen, 1992, 1996; Saracevic 1996, 1997; Belkin, 1993, 1996). O'Connor's (1978) interaction model addresses the statistical product of how a person and situation characteristic can be defined as a unique contributor to explain variability (or effect size) in an information seeking decision making environment. His model depicts there are endless interactions between person characteristics (as internal stimuli) and situation characteristics (as external stimuli) as a person engages in information seeking to make a decision at each step of the search process. O'Connor's (1978) interaction was achieved by injecting newly created interaction variables in nonlinear regression models. Later studies by others would, from time to time, report statistically significant interactions using analysis of variance (ANOVA) models. These statistical artifacts confounded the models since they obscured the importance of significant main effects hypothesized in the original design. It is not productive to visit these ANOVA models but it is noteworthy that they affirm that interactions are occurring across variable factors in IR research.

The underlying basis for a user interaction can be partly traced to the work of Ingwersen (1992, 1996, and 1999) who introduced a model which viewed such interaction as a process of users' cognition. Based on Belkin et al's (1982a, 1982b) ASK concept, this interaction relied on the position that users have an inadequate knowledge of their own needs and the systems' capabilities. The interaction of the person with the system was seen as intertwined over time and a new or revised model was introduced by Belkin (1996) & Belkin et al (1995) labeled the "Episode interaction model." This episode model posited there are a series of different interactions occurring between the user and system as users seek information with different seeking strategy and the system provides a different and changing response to revised queries. This model includes situational and environmental factors, such as users' goals, tasks, knowledge, intentions, the history of episodes and problem in addition to users' cognition. This model is related to the current study which draws upon several of the more salient variables uncovered in the episode line of research. Tasks become particularly important as users interact with IR systems. These models were tested in English language environments.

Kuhlthau developed the ISP model which is based on a series of empirical studies using various research methodologies within a longitudinal design to investigate high

school students' and college students' information seeking processes (Kuhlthau, 1983, 1988a, 1988b, 1991, 1993a, 2004). The model holds that people search for and use information differently depending on the stage of the process. The processes are presented in six stages describing the user's thoughts, feelings and actions within each stage. The six stages include: task initiation, topic selection, pre-focus exploration, focus formulation, data collection, and closure. At each stage, the "thoughts", "feelings," and "actions" are different as the person moves to the next stage reflecting how change occurs in the information seeking process. Kuhlthau also identifies four criteria that might affect information seekers' search processes to choose information: task, time, interest, and availability of the information (1993a, p. 39). These are critical criteria used to judge relevance when information seekers conduct their information search. Kuhlthau's ISP is initiated by 'uncertainty' or a lack of understanding which will be changed gradually as the information seeker obtains information, reduces uncertainty, and resolves problems (Kuhlthau, 1991, 1993a). This research also informs the current study since it brings to the fore the need to assess user's perceptions of the availability of information. Clearly, in a multilanguage search, users might expect that the information needed is being blocked by technical aspects of transliteration. The Kuhlthau model also emphasizes the importance of task and this is incorporated into the basic experimental design of the current research study.

Saracevic (1996, 1997) created a comprehensive model called the stratified IR model which proposes that there is a sequence of processes (interactions) occurring in several connected levels or strata between user and system through an interface as a surface. The levels or strata represent each different element, such as query (characteristics), cognitive (knowledge, structure...), affective (intent...) and situational (tasks...) on the user side; and, engineering (hardware, capacities), processing (software, algorithms...), and content (information resources, representations...) on the system side. As the information is processed, it gets adapted through interactions and further modified by such variables as feedback from the system. Thus, the user functions at various levels or strata as does the system and this interaction then leads to even more interaction. The present investigation uses a part of this model in assessing how queries are modified as searches progress. The focus here is on a quantitative assessment of such change and its impact on search outcomes.

IR studies have not been located which propose interactive models to explain how non-Roman alphabet users' retrieve information from multilanguage systems. The models highlighted above have emphasized monolingual user' information retrieval processes and it is interesting to apply parts of these conceptual models to an explanation of how users seek non-Roman alphabet information.

2.4.2. User information search behavior in interactive IR systems

2.4.2.1. Studies in monolingual IR systems

A query is a representation of a user's information need, which consists of search terms and of possible operators connecting them (Vakkari et al., 2003, p. 449). There have been many studies that examine users' search behaviors either with real users or with transaction log data analysis. With real users, the study designs include a number of different user groups broken out by various characteristics: by their age (children: Bilal, 2000; Abbas, 2005); by professional status (historians: Bates, Wilde, & Siegfried, 1993; as graduate students: Shaw, 1995; Rieh & Rieh 2006; Jeong 2004); by levels of searching expertise (Hsieh-Yee, 1993); by domain knowledge (Hsih-Yee, 1993; Marchionini et al, 1993); as students (Rieh & Xie, 2001, 2006; Spink & Jansen, 2004; Wang, Berry, & Yang, 2003). Hsieh-Yee (1993) investigated the effects of subject knowledge and search experience on novices' and experienced searchers' use of search tactics in online searches using the ERIC database on the DIALOG system.

The above studies have employed a variety of methods: query analysis (Jansen & Spink, 2006; Rieh & Xie, 2001; Spink & Jansen, 2004; Wang, Berry, & Yang, 2003; Jansen, Spink, & Saracevic, 2000; Hsieh-Yee, 1993; Beitzel et al. 2004); laboratory experiments with transaction logs analysis; on-site observation (Xie, 2002; Kim, 2002; Wang, Hawk, & Tenopir, 2000; Hsieh-Yee, 1993; Yi et al, 2005); and, observation and interviews in a natural settings (Kim, 2005; Rieh & Rieh 2006; Rieh, 2004; Xie, 2002).

There are a number of studies which have used transaction log data from search engines and OPACs (Jansen & Spink, 2006; Jansen, Spink, & Saracevic, 2000; Borgman et al., 1996; Spink & Saracevic, 1997). Jansen, Spink, and Saracevic (2000) found 67% of Excite.com search engine users submit their queries using a single word. These various methodological approaches point to the need to use multiple methods to study multilanguage information retrieval. As a minimum, a survey and an experimental setting appear appropriate to uncover the processes and experiences of those seeking information in languages other than their own while taking into account different purposes for searches, different tasks, and different language abilities.

Search behavior research in interactive IR systems has mostly focused on users' interaction with a system by analyzing basic search strategies (Chen & Dhar, 1991; Xie,

2002); tactics (Bates 1979, 1990); moves from one query to another (Fidel, 1985; Bates, 1999); and search processes (Kuhlthau, 1991, 1993; Ellis, 1989). Interesting to note here is that various users may exhibit particular styles when searching for information. In fact, the collection of search data in a research setting exposes a number of these interesting variables where data naturally emerges or can be collected in a straightforward way. Various variables from such analyses include: search times (Spink & Saracevic, 1993; Fidel, 1991; Wildemuth, 2006); the number of search terms and the overlap among search terms (Saracevic et al, 1988a, 1988b, 1988c; Jansen et al., 2000); the number of search cycles (Wildemuth, 2006); queries which emerge from natural languages or are derived from controlled vocabulary (Hsieh-Yee, 1993; Wildemuth, 2006; Fidel, 1991). Although these studies do not funnel neatly to a single conclusion, they show the diverse array of concerns important in making Web access compatible with users' needs and information styles.

2.4.2.2. Studies in CLIR systems

F.C. Gey et al. (2005) proposed that the future of CLIR research should bring more attention to end-user issues such as results presentation and multilingual question answering (p. 424). They also pointed out users of CLIR systems are mostly looking for and retrieving information in languages in which they have little or no competence. So the system has to provide help when a user formulates a query and attempts to interpret the result. From such findings, several researchers have recognized the importance of user – system interaction and the value of building initial models based on real user participation (Petrelli et al., 2008 & López-ostenero et al. 2008).

Such models have implications for developing theory which might guide future research. On the one hand, if theories are derived from empirical generalizations, then the theories may be based on monolingual IR systems and not take into account the value of a more complex information system. On the other hand, studies which begin with theory prior to the establishment of an empirical base may overlook the importance of simple concepts and obvious variables which influence user-system interaction using a multilanguage IR system.

This current study attempts to do both: propose a robust theory while attending to the importance of empirical observations. It is assumed in this study that theory develops from hypotheses and not solely from empirical generalizations but that there is interplay between these two. Such an approach will drive data analysis in the current study where attempts will be made to derive models from collected data.

There was a recent question and answering (QA) user study done as a component of an interactive CLIR. López-ostenero et al. (2008) compared experimental subjects' search accuracy when they searched for cross-language information, English vs. Spanish. This study included machine translation versus monolingual searching. In this experiment, the researchers used Systran Professional 3.0 by the iCLEF (Cross-Language Evaluation Form) organization as the machine translation (MT) software, and the data set used was the iCLEF 2004 question set in English. The study's results show that cross language searching was not incrementally harder than monolingual searching except that cross language searching took more search time than monolingual searching. Since the languages they compared was English with Spanish, it can be assumed the MT does better job than other language pairs, especially if matching a Roman alphabet language with a non-Roman alphabet language.

Although this research was done using a small number of subjects (n=16) with a limited data set, it is noteworthy that the researchers attempted to measure and observe real users in an interactive CLIR environment. It needs to be pointed out that studying language pairing with an IR system might not be a productive avenue for research since even a small number of languages can produce a large number of unique pairings (for example, 10 languages would produce 45 pairs of languages to be studied). The approach adopted for this study addresses language as an aggregate and it compares Roman with non-Roman alphabets while accounting for task, experience, and user knowledge of languages.

So far, some researchers have explored a user interactive approach with a CLIR system. For example, the QUILT system provided a display of the Spanish translation of English query terms (Davis & Ogden, 1997); and, the Arctos system offered a browserbased interface with which to enter English queries. In the Arctos system, the user can interactively improve the query translation using links to on-line bilingual translation resources (Ogden et al., 1999). The Keizai system allows users to go from Japanese and Korean Web data to displays with English summaries of the top ranking documents (Ogden & Davis, 2000). Other systems allow query expansion, (Ballesteros & Croft, 1998). The C*ST*RD is an interactive information access system using space visualization, which can make it possible to open a document or content according to the end-users' selections. There was also a study of the Hindi language for English speakers which were tested (Leuski et al. 2003). All of these novel approaches have merit but it may be likely that they will be so specific to particular users, situations, and systems that the generality of their results might be unduly restricted. The study proposed here will make an attempt to provide an information environment and a user experience commonly found in general IR users studies while injecting major language differences.

CLIR system development is now actively studied based on users' needs although its current development stage still focuses on the physical system itself. However, because of its unique situation covering different languages, the system design should give equal attention to constructing algorithms and to defining user needs. Attention to each language's characteristics, representations, and specific needs should also be explored as they may apply to particular intelligent interfaces. All these factors are essential or necessary prerequisites for structuring system components and features to make them more user-centered. Also, it is hoped that any future system could produce more efficient representations of responses to users' queries within the framework of a culturally sensitive and helpful interface. Such a goal is actually beyond the scope of this investigation but its overall premise is valuable and should guide future studies in the CLIR area.

2.4.3. Language preference in query formulation

In this research, users' choice of languages will be focused on when they search non-Roman alphabets, particularly searching for CJK language information. There has been an assumption that multilingual users are likely to submit their search queries in their first language since they may not be able to express their information needs well enough in non-native languages (Ogden & Davis, 2000). Yet, an early study by Petrelli et al. (2004) found that users chose the most appropriate language for their task and that was not necessarily their native language.

Rieh and Rieh's study (2005), where they analyzed bilingual researchers' behaviors, perceptions, and preferences when they use IR systems for research, reconfirmed Ogden and Davis' (2000) study. The subjects who were Korean science and engineering scholars tended to chose a language that can represent their information needs most accurately rather than always selecting their native language or English (p. 255-260). That is, user choice of language is dependent upon types of tasks, rather than familiarity with the language. These studies are useful in mapping language construction of original queries, even though the studies employed limited sample sizes, and, at times, arrived at contradictory findings. The salient result is that there is a critical need to develop many more user studies in this area to insure that the information search processing involving different languages is better understood. Based on the above findings, the current study will examine search performance, as reported by the users, while noting if the query language selected is the subject's native language or the language of the database for a particular search.

Query changes are also important in traditional IR research. It might be assumed that the review of changes to queries would produce a series of logical patterns which could emerge from the mind of the user. In the current study, searchers' query formulations, especially in their language choices and their examination of bibliographic records will focus on individuals' assessments made during the search while seeking non-English information. This study will also note if there are different search behaviors by subjects' with different language backgrounds. In addition, subjects' language knowledge will be assessed. The subjects' use of English vs. non-English searching will also be explored.

The specific variables relating to query — as it has been studied in research whose designs are similar to this study — include: number of terms subjects' use (Saracevic et al, 1988c ; Jansen et al., 2000), use of Boolean logic (Marchionini, 1988; Sewell & Teitelbaum, 1986); Trzebiatowski, 1984), and use of advanced search functions. These analyses will focus on comparing English with non-English searching to see if there are differences between the constructions of the two searches. This study does not directly cover other search behaviors, such as tactics (Bates 1979), moves (Fidel, 1985), strategies (Chen & Dhar, 1991; Xie, 2002), or search process (Kuhlthau, 1991, 1993; Ellis, 1989) although variables related to these are expected to be revealed as experimental subjects are observed. Nonetheless, the main approach here is depicted in Figure 1 where search behaviors are assessed in an experimental setting and from self-reports on surveys.

2.5. Bibliographic record's evaluation and its selection

2.5.1. Relevance in general

Since information science's beginnings, information retrieval research has focused on how people seek information and how they judge its relevance. *Relevance* cannot be defined simply by saying 'how pertinent is the information retrieved from a certain system which can serve users' particular needs.' Relevance judgments occur as individuals evaluate surrogate language (such as index terms and abstracts) and how they evaluate and use the documents retrieved. System-centered approaches dominated the early study of these phenomena and they used such metrics as precision and recall to assess the likelihood of retrieving relevant documents from a database.

Later approaches were more user-centered and extended the situation from a system focus to include cognitive variables which might account for explaining users' relevance evaluations (Saracevic, 1975; O'Connor, 1978; Ingwersen, 1982, 1992, 1996; Belkin & Vickery, 1985; Borgman, 1984; Dervin & Nilan, 1986; Belkin, 1980; DeMay, 1977; Wilson, 1981). Relevance as a subject of study matured as different research investigations accrued over time. Relevance evolved to encompass the wider information environment, the importance of person and task, and the impact of reforming queries as humans and systems interacted.

Such concerns expanded to include recognition of important situational elements in the retrieval environment: the organization (social domain), the individual, and the cognitive complexities of a human working for an organization (Schamber et al., 1990; Schamber, 1991; Barry, 1994; Cosijn and Ingwersen, 2000). Barry and Schamber (1998) conducted similar research to assess end users' relevance criteria. Both used small sample experiments making subjects conduct a search followed by an interview with each searcher. The interview transcription was analyzed using content analysis. The content categories were summarized into seven classes: information content of documents (accuracy, in-depth information, etc), source of documents, physical entity, other information sources within the environment, users' situation, users' beliefs and preferences, users' previous experiences or background.

Saracevic (1999) emphasizes a key point by stating, "The issue is how to deliver and incorporate the desired design features that will improve systems orientations toward uses, integrate them with systems features, and use advantages provided by both humans and technology" (p. 1058). After this conceptual approach emerged, there have been many research trials and studies to combine both a "systems-centered" and "human-centered" integration rather than separate developments of these related aspects.

There are several studies which define relevance and its criteria. Saracevic (1975, 1996 & 1997) provides a comprehensive view of relevance noting its component characteristics, yet acknowledging its elusive nature. He emphasizes a need to see the whole process of information searching — almost a holistic model — and this is reinforced with his stratified IR model. He defined relevance as manifestations of relevance (1996 & 1997): 'system or algorithmic relevance, topical or subject relevance, cognitive relevance or pertinence, situational relevance or utility, and motivational or affective relevance.' The importance of this work is its position that there cannot be a separation of the components of the information search: all elements are connected and interacting with each other. This informs the current study proposed here which posits a staged model where prior behaviors influence later ones as each component of the search accounts for variability in later relevance decisions (see Figure 1).

Schamber (1991) also presented ten summary-level categories from her study: accuracy, currency, specificity, geographic proximity (since the subjects were from weather related work), reliability, accessibility, verifiability, clarity, dynamism and presentation quality. Thus, as Cosijn and Ingwersen (2000) argued, the relevance judgments are representations of information objects and these results are from "manifestations of socio-cognitive relevance" (p.546, 549). Hyldegård (2004) also argued it is important to consider users' social and collaborative dimension in their information seeking. It is because the system should support various user roles, needs and types of interactions during a problem solving process (p. 277). Thus, person characteristics might be assumed to explain a portion of the variability as individuals seek information in various languages and arrive at relevance decisions regarding their search.

There are other views which can be derived from a domain analysis perspective which recognizes the organization, the individual, and the cognitive complexities of a human working for an organization or driven by a particular need that relies upon the performance of an IR system. The cognitive view in IR has been criticized in some areas as a "lack of realism", "lack of theory integration", (Ingwersen, 1999) and "isolate the individual from the social, cultural, and historical influences" (Hjørland, 1997) (Cited in Jacob & Shaw, 1998, p.140). De Mey (1980, 1982) indicates that this cognitive view places too much emphasis on the organizing activities of the 'individual' – "model of his world", and pointed out it should follow on the "*common cognitive process in individuals functioning within certain patterns of social organization*" (Cited in Jacob & Shaw, 1998, p. 136-137). Yet, what is the perceived in the eye of the searcher can be said to define relevance, especially as individuals move across languages and alphabets.

Marc De Mey (1982) pointed out that Information Science should be aware of the "common cognitive process in individuals functioning within certain patterns of social organization" (p. xvi). Blair (1990) and Shera (1965) also argued in information retrieval system design — with special attention to representation of texts in a system — that IR should encompass a thorough examination of language usage, which is socially constructed within social processes. Jacob and Shaw (1998) emphasize that we need to investigate "the linguistics, communicative, and organizational aspect of representation from a multiplicity of sociocognitive perspectives and within the full range of discourse

domains and knowledge communities" to obtain thorough knowledge representation and organization in information retrieval (p. 170-171). Hjørland & Albrechtsen (1995) also proposed that a new approach, called "domain-analysis," is needed to understand information in information science. In Hjørland's recent paper (2002), this socio-cognitive perspective is well represented:

"This view changes the focus of IS from individuals (or computers) to the social, cultural, and scientific world. One important implication is that the relevant cognitive structures are of a historical rather than of a physiological nature. In developing this view I found support inside psychology from *the Cultural-Historical Approach* associated with names like John Dewey, L.S. Vygotsky, and A.N. Leontiev, also known as *Activity Theory* and *the Sociocognitive View* ⁴(p. 258)."

This sociocognitive concern then needs to address how the interactions might occur. Shera (1965) emphasized that an optimal level of retrieval can be achieved only when information professionals develop and implement representational systems that establish congruence between the cognitive organization imposed on knowledge by individuals and the conceptual organization imposed on documents by the information

⁴ "Socio-cognitive" theory originated from "socio-cultural theory" by soviet psychologist Lev Semenovich Vygotsky. According to Wertsch, who introduces and interprets Vygotsky's major theories, Vygotsky's socio-cultural concept can be explained as "higher mental processes in the individual have their origin in social processes" (Wertsch, 1985a, p.14). That is, human intelligence originates in our society or culture rather than from an individual's intrapersonal self. This theory indicates that an individual's cognition can be formulated through continuous interaction with a social environment that has inherent historical and cultural characteristics. Related to the social interaction and mental processes, Vygotsky regards the human language as the most influential medium among other tools (Wertsch, 1985a, p.15).

Vygotsky's socio-cultural theory has been used by many scholars from several different fields – such as psychology, education, sociology, linguistics, cognitive science, philosophy, and information science – as a fundamental framework for their studies (Luria, 1979; Wertsch, 1985a, 1985b; Grabois, 1999; Echabe, & Castro, 1998; Hjørland, B., 2000). In particular, this theory has been widely studied within the cognitive-linguistic areas. For example, Grabois (1999) attempted to bridge "socio-cultural theory" and cognitive linguistics. The author conducted an experiment to see "how the emotions can influence our knowledge of other concepts, and how this can be variable across culture" by using "word association" methodology. The experiment was conducted with different language groups (i.e. native Spanish speakers, expert speakers of Spanish, Spanish learners, foreign language Spanish learners, native English speakers) using emotional words (i.e. "love," "happiness," "fear" and death").

specialists (Cited in Jacob & Shaw, 1998, p.134).

To process this, especially for multilingual users, a more efficient and helpful interface may be a crucial component in improving the search process because it can be expected that most users do not have a perfect knowledge of languages different from the user's native language. To offer a reasonable interface, when users process their information search, the system might be expected to offer various options to approach and obtain the target queries or information — regardless of query or document language. For example, this suggests that broader terms or common words be used within a particular language domain and it presupposes that the system has the ability to offer synonyms with easy access to cross-referencing.

CLIR systems are more complex than monolingual systems and the CLIR environment will thrust users into more complicated IR situations. This, in turn, may force the reexamination of assumptions and findings which come from user studies of single language retrieval systems. Information needs, language uses, and cultural differences may need to be studied anew within a CLIR environment. Correspondingly, there have not been many studies examining relevance in a CLIR environment. Arnold et al. (1997) argued that the operational evaluations should take into account the end-user's subjective measures of translation adequacy (such as intelligibility, accuracy), coverage of linguistic phenomena (such as via creation of standard tests). It is recognized that it is difficult to model a single user's information search behaviors with a monolingual IR system; it may be even more difficult to extend this to a CLIR system.

Hansen & Karlgren (2005) used a machine translation system and asked Swedish subjects to judge English documents. Orengo & Huyck (2006) asked Portuguese users to judge the relevance of English documents. These studies affirmed that research involving relevance in a CLIR environment is within a more complicated situation than those examining monolingual information retrieval users. Crossing language and even alphabet barriers becomes important variables influencing articulation of the information need and its expression through the creation of search queries and examination of retrieved bibliographic records and even intentions to use documents. Language differences can also embed cultural differences and this further complicates the search process and the evaluation of results.

After all, the information environment needs to encompass an understanding of the users' within the culture and society if the system is to be compatible with the complete user who is, after all, a complete person with a myriad of variables accounting for thought and behavior. To process this, it is crucial to tap into a comprehensive users' analysis of queries within the users' culture and context and, equally importantly, to have the system structure index terms (representations) extracted from documents of the system's collection to create an interactive IR (Information Retrieval)/CLIR (Cross Language Information Retrieval) environment. The terms should be created considering the users' perspectives based on their social and domain knowledge, so that the system can yield acceptable and understandable results to users.

As applied to this study, to set forth these concepts in the global communication of bibliographic data implies an understanding of different language backgrounds of users. This would include their perspectives and it would be regarded as a fundamental beginning when creating a user-efficient system in global information retrieval. That is, an examination of the user, the task, and the person's use of a particular system could provide clues on what might constitute a more efficient interface and a better representation of a document. It would do this while acknowledging the importance of language-specific meaning and the users' perspectives within a social organization. It has not been found that this particular approach has been incorporated in IR systems, nor have these concepts been expressly posited as important in furthering understanding of non-Roman alphabet information users.

2.5.2. Intention to use information based on bibliographic record

Taylor (2000) identified the three main functions of bibliographic tools: (1) identifying or finding; (2) collocating or gathering (keeping the same works by the author or on the same subject together); and, (3) evaluating or selecting. Relevance occurs in every step of the information process using an IR system, such as selecting a database to search, articulating the user's information needs as search queries, reviewing the retrieved bibliographic records, assessing the search results as they are perceived to correspond to the users' topical knowledge, system knowledge, background, information need, goal, task, priority of the task, and cognitive abilities. In this section, the relevance concept will apply to the evaluation and selection of bibliographic records with user indications on whether they will intend to review the real documents retrieved.

Tagliacozzo and Kohen (1970) noted that cataloging information can provide at least two levels of decision making prior to evaluating actual library materials. The first level is the user's decisions on the catalog filing term, while the second level addresses an assessment of the catalog record's bibliographic description of an item. The former concern addresses the user's query formulation and how it matches a representation of a document, while the latter involves decision making regarding intention to use the retrieved and selected information. This approach was applied in O'Connor's (1978) research where he tried to measured the relevance of "intention to use" when subjects examined a simulated library catalog.

Similar to this study, Cooper and Chen (2001) conducted a study where they predicted the relevance of a library catalog search based on library search log data where four actions were indicated: print, mail, download and save. This study takes for granted that users would review the documents when employing different aspects of relevance decision-making for certain intended use indicators (such as download or email).

Park (1992) conducted a user study to identify factors which affect users' relevance decisions when they examine document representations. She categorized this according to three factors: internal context (users' prior experiences or perceptions such as level of subject knowledge, research experience, and education level...), external context (system search and current research, such as perception of the search, perception about the availability of information, purpose of the search, priority of information need...) and problem context (motivations about the intended use of a citation, such as obtaining definitions, background information, and, methodological framework for the problem ...). She also explored how users interpret the information from the titles, bibliographic citations and abstracts based on users' comments. This approach is similar to a portion of the investigation proposed here.

There have been few studies dealing with bibliographic information evaluation in terms of interface design (Cherry et al., 2006; Grefenstette, 1998; Chan, 1995; Cherry & Cox, 1996; Thomas, 1997). Grefenstette (1998) examined the features of a bibliographic record's content that users want to have when they use a standard or classical information retrieval system. He found that users want to see the bibliographic description include or give direct access to abstracts, paragraphs, and then articles. This notion can be applicable to the same expectations that CLIR users may have where the bibliographic description is expected to provide full and detailed information including an abstract, a summary of the document, as well as basic description, (e.g., title, authors), in their native languages and/or with English. Thomas (1997) examined users' preferences for the bibliographic information display. He conducted an experiment asking real users about four different bibliographic information records, such as with or without data labels and with or without detailed data entry. The results indicated that users preferred brief information rather than a full record of all details. These studies did not look at the availability of information in different languages nor did they inquire about the need for an abstract.

2.6. Summary of Literature review

In the above literature review, four different but related subjects were presented: access to non-Roman alphabet information, user information seeking behaviors, user's information search behaviors, and bibliographic record selection and evaluation. As seen in the literature reviewed above, there have been many issues raised regarding monolingual issues and even more issues addressed to studying users of CLIR systems. There remains a dearth of studies dealing with MLIR/CLIR research which focuses on non-English information users.

Chapter 3: Theoretical proposition and research questions

3.1. Theoretical background

The theoretical premises underlying this study rest on the foundation established in the literature review where productive findings from related research are incorporated into a model which proposes to explain user assessments of and behavior in a multilanguage information retrieval environment. As noted, special attention is given to issues which become evident when users cross from one language alphabet to another. The theory, then, is best depicted by the model prepared in Figure 1 provided in the first chapter. It is expected that the findings from the survey and experiment launched in this study will inform the model and result in its modification for future research.

Theory can be thought of as the specification of the relationships among concepts. In turn, concepts are considered as abstract classification terms from which observable, empirical variables are derived. Furthermore, it is assumed here that theories differ by how they account for variability and the measurement of such variability, or effect size in a quantitative model, represents a non-trivial consideration. This Chapter attempts to set forth the specific conceptual framework which provides the basis for the design aspects of this research to include the selection of variables.

From earlier chapters it can be seen that the search process confronting users of multilanguage IR systems involves the following: the user, the task, the query, the search term opportunities including the index languages available, the content of the data base, the ranking of results, the display of records and documents, and the availability to modify the search to achieve different results. Overlay onto to this the complexity of thinking of the task in one language, constructing a query in another language, and then obtaining results in a multitude of languages. This information environment is further complicated for the user when Roman and non-Roman alphabets are intermixed in the search and retrieval processes.

The information environment with user and system can be thought of as a social situation. The personal characteristics of the user and that person's assessment of task and purpose can be framed within a person centered theoretical perspective. The person behaving in the information environment can be seen to represent instances of user/system interaction. Hence, the current study can be classified as relying on a person centered viewpoint where language considerations have been appended to the theoretical contributions of the productive IR studies summarized in Chapter Two.

3.2. Theoretical proposition

Based on previous experimental models and the literature review reported in the last chapter, a tentative explanatory model was constructed (Figure 3.1. explanatory model). Major influences in variable selection and definition have come from a number of related studies. Some of these will now be discussed in more detail.

O'Connor's (1978) interaction model attempted to prove statistically how much variability is explained by person and situation characteristics when users make a decision in an information seeking environment (relevance judgment as an intention to use certain information). His model emphasizes interactions between person and situation characteristics which are posited to make separate contributions to explaining behaviors as users proceed through an information seeking process. He also indicated that the

statistical interactions (mathematically derived) create their own variables which contribute unique, non-redundant contributions to the explained relevance variability (i.e., effect size) using numerous curvilinear transformations of separate person and situation variables. The experimental approach used was to create an information search situation, assess person variables and situation expectations, explore non-linear relationships, add interaction variables, and stage the explanatory model to track with the sequence of search processes conducted by the user. The descriptive model was then tested with new data to assess the accuracy of the original coefficient weights. In all, the staged models accounted for 25% to 40% of the relevance variability and a discriminant model had high prediction accuracy for system preferences. O'Connor's research is interesting but it is less relevant in this initial study of user behavior in a multilanguage environment. It is summarized here because its interaction variables accounted for nontrivial relevance variance although its impact is primarily as a methodological contribution. After this current study is completed, it might be expected that future research in CLIR user studies might explore the potential for effect size contributions made by creating new variables as statistical interactions.

Kuhlthau's ISP model provides clear insights into how users' seek information depending on the stage of the information process. The model also encompasses how users' emotions are altered during their process of searching information. Kuhlthau's ISP is initiated by 'uncertainty' and it is assumed that non-Roman alphabet information users might have this uncertainty and emotive state when they access and use current IR systems. The current study has been influenced by the Kuhlthau model although the methodology used here is biased toward a quantitative approach to accounting for user search behaviors. In all honesty, the current study might be viewed as capitalizing on low hanging fruit in its selection of known IR variables while adding to these with a complex language environment.

Saracevic's relevance theory and stratified IR model also provides insights which can view the information process as a whole and, at the same time, assess the importance of each element in the stratum as user and system interact. Each stratum has a unique role in information processing and interactions occur in and between each other. So the elements are intertwined with each other and it becomes a condition and then effect of the relevance decision process. Considered here is the notion that the condition, necessary and/or sufficient, might be used to test elements in the stratified model and also add language complexity to the overall search process. Hence, the current study draws from the Saracevic's research considerations regarding query construction, cognitive structure, affective intent, and situational tasks.

The proposition underlying this study is that "access and use of non-Roman alphabet language information is a function of 1) person's characteristics (includes age, education, native language, and work experience); 2) person's experiential knowledge (includes topic, system, language); 3) situation (includes task, topic, system coverage, system retrieval function, and availability of the system); 4) query construction (includes first language choice, search time, number of query changes, difficulty index); 5) IR system preference (includes system choice and users' termination of their search); 6) bibliographic record understanding (includes understanding Romanization and contents); 7) users' satisfaction with the system and the search; 8) system efficiency for non-Roman alphabet information; and, 9) relevance intent to use the retrieved information (Figure 2).





The model proposes assessing the influences of person characteristics and knowledge experiences as individuals embark on selection of retrieval system, engage in query construction, and evaluate and indicate expected use of retrieved information for non-English and English languages. An emphasis of the study is its exploration of how individuals might use Chinese, Japanese, Korean (CJK), and English as they search for scholarly and other information. As chapter two notes, CJK languages are spoken by a significant proportion of the world's population and these languages represent a growing influence in the WWW and the Internet. The model depicted above is expected to be modified after all data analysis is done.

3.3. Research Questions

The literature reviewed and the theory considered suggests five research questions related to CLIR system users' information seeking behaviors:

- 1 What are the patterns of non-Roman alphabet users' information seeking and use behaviors?
 - 1.1 How do users use current information retrieval systems searching for non-Roman alphabet information, especially via online DBs, OPACs and web search engines?
 - 1.2 What are the users' needs and expectations when accessing non-Roman alphabet information using such IR systems?
 - 1.3 What are their language choices and considerations when they construct queries?
 - 1.4 How do users' language knowledge and background affect their non-Roman alphabet information seeking and use, if at all?
 - 1.5 Are their user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine?
- 2 What issues are present when individuals search non-English information via information systems?
 - 2.1 What kinds of issues and limitations exist when searching non-English information?
 - 2.2 How do those factors affect the access to information contained in such systems?
- 3 How do bibliographic records facilitate or hinder the understanding of bibliographic and retrieved information?
- 4 What explains non-English information users' relevance certainty judgments (intention to use retrieved information)?
- 5 What would constitute an explanatory model for users searching for non-English information? How does this model explain influences of user characteristics, language, topic, and task in their query formulation and relevance evaluations of bibliographic information?

3.4. Premises/hypotheses for each research question

Table 2.	Premises	hypotheses	for each	research	question
					1

Research questions	Premises/hypotheses		
1. What are the patterns of and non-Roman	R 1.1 & R 1.2: Users' search behaviors and special needs when using multi-		
alphabet information seeking and use	language systems will be identified.		
behaviors?			
1.1.How do users use current information retrieval systems searching for non-Roman alphabet information, especially via online	H1.3: Based on background and experience, users will input (or prefer) to state queries in their native language rather than Romanized word or English.Assumes that users prefer to search using their native language.		
1.2. What are the users' needs and expectations when accessing non-English information when using such systems? 1.3. What are their language choices and considerations when they construct their queries?	 H1.4: There are statistically significant differences in users' IR system use and search behavior by users' background and their experiential knowledge. Expect to find different behaviors by users' background (language background, professional role) and their experience in system use, language, or topic knowledge. 		
 1.4. How do users' language knowledge and background affect their non-Roman alphabet information seeking and use, if at all? 1.5. Are user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine? 	 H1.5: There are statistically significant differences in users' IR system use and search behavior for non-English information when compared to English language searching. Expected to find different search behaviors (number of terms used, number of queries changed, search time, Boolean logic use, language choice, and advanced tools used) by English vs. non-English searching Expect to find different evaluation score for users when searching non-English information compared to English language searching. 		
2. What issues are present when individuals	R2.1: Users' issues when using multi-language systems will be identified.		
search non-English information via online DBs,			
OPACs and web search engine?	H2.2: Expect to find there is a predetermination whether a specific OPAC or		
2.1. What kinds of issues and limitations exist when searching non-English	database will be used or not by its language coverage or CLIR function availability. Assume that users will devalue or even abandon their search if the		

information?	system does not have their target language collection.
2.2. How do those factors affect access to information contained in such systems?	H2.2.1: Users will avoid databases if they perceive that the system lacks efficient and comprehensive language coverage.H2.2.2: Users' searches will significantly shorten their search time using online database for journals compared to WorldCat or Google search.
	H2.2.3: Users within their area of expertise will have better knowledge about scholarly academic journals written in English than journals written in non-English.
3. How do bibliographic records facilitate or	R3: What kinds of bibliographic records are helpful to understand bibliographic
hinder the understanding of bibliographic and	and retrieved information? Assumes that users need full records which include
retrieved information?	English abstract and translations of key entities such as title, descriptors and vernacular language.
	H3.1: There will be statistically significant differences in users' understanding levels when assessing full records which include English abstract, English translation and vernacular language; furthermore, these records will get higher scores than other record displays.
	H3.2: There will be statistically significant differences in bibliographic understanding levels for English searching vs. non-English searching.
	H3.3: There will be statistically significant differences in bibliographic understanding levels by researchers compared to library professionals.
4. What explains non-English information	R4: Different relevance criteria for non-English information users are expected
users' relevance certainty judgments (intention	when compared to monolingual English information seekers.
to use retrieved information)?	

	H4: There is a statistically significant difference in relevance certainty judgments for searches seeking non-English information compared to searches seeking English information.
5. What would constitute an explanatory model for users searching for non-English information? How does this model explain influences of user characteristics, language, topic, and task in their query formulation and relevance evaluations of bibliographic information?	R5: An explanatory model of searching for non-English information will be constructed from the data analyses and it will redefine the original, hypothesized model.

Chapter 4: Methodology and data collection

4.1. Overview

As introduced above, the design of this study includes two major components: an experiment and a survey. The multi-language experiment used observation and interview as it examined individuals' searching patterns for different topics using three databases which provide query access and retrieval using different languages. Thirty-two subjects participated in this experiment requiring use of Multilanguage information sources. The experiment rotated tasks, queries, and three different databases: WorldCat, EBSCOhost, and Google Language Tools.

The parallel study was an online survey which explored issues that arise when individuals search for non-English information. A total of 204 respondents participated from academic researchers, library personnel, and the general public who were solicited through a network of colleagues. Of special interest is access to and understanding of Chinese, Japanese, and Korean information in online catalogs, scholarly databases, and web portals.

The design of both studies emphasized the creation of explanatory models using quantitative statistical data analysis. Nonetheless, qualitative data was collected in selected areas with an intent to serve as a confirmatory approach to affirm particular quantitative results. Transcriptions of the interviews, observations, questionnaires, think-aloud protocol, survey data, and results assessments were primarily coded to build statistical models. Subjects' comments and other related qualitative data were injected, when appropriate, to give a more expansive meaning to the constructed analytic models.

For qualitative analyses, the descriptive data from both experiment (think aloud comments, observation, open-ended questions) and survey were coded using portions of a grounded theory approach (Strauss & Corbin, 1990). The data was coded without any initial regulation at first and then the coding proceeded to reveal enriched aspects of the user's experience when searching multilanguage databases for different tasks.

Experimental data were analyzed using case study scenarios with appropriate content analyses and these are augmented by analysis of variance models (ANOVA) to assess differences across users and languages. Extensions of the ANOVA models using the generalized linear model and other regression routines are also employed in the construction of the explanatory models. The questionnaire data was tested against specific hypotheses using regression analyses and ANOVA models (such as t-tests when appropriate) to construct descriptive and inferential models.

4. 2. Experiment with observation and interview

4.2.1. Overview of the experiment

A total of 32 individuals from different language cultures searched randomly selected topics using their own information seeking strategies with three different systems (WorldCat, EBSCOHost, and Google) yielding a total of 288 searches. These individuals' native languages were Chinese, Japanese, Korean and English. The results of this experiment affirmed the original theoretical model and it also added new information regarding how non-Roman alphabet information users regard access to and use of scholarly and other information searchable in different languages.

Major findings constitute how certain variables influence individuals as they

progress through a search from query formation to intention to use information. It had been proposed that personal characteristic variables and search environment variables (for example, query construction and database used) would be of importance in explaining the early factors involved in searching. Such a model unfolds as it progresses through the various search stages with earlier dependent variables (DV) being used in later sequences as independent variables (IV). For example, if a person's experiential knowledge of topic (IV) can explain query changes (DV), then — at the next stage these two variables, knowledge of topic (IV) and query changes (now considered an IV) can be used to explain users' satisfaction with system search results (DV).

The user's language choices and modifications of their queries to achieve optimum retrieval results are a particular concern of this experiment. A query can be considered as a representation of a user's information need, which consists of search terms and of possible operators connecting them (Vakkari et al., 2003, p. 449). This study will analyze search terms to include subjects' query formulation, and repeated runs of queries. By reviewing changes to queries, it would be expected that logical patterns may be evident in the mind of the user.

Think-aloud protocols were employed to explore subjects' language use with each bibliographic system, their interaction with the system in terms of forming and reforming their search queries, their use of search terms to initiate searches, and their evaluation of the retrieved bibliographic records. It is important to note that these protocols were directed at confirming results also obtained via data collection instruments provided to the subjects. These instruments, appended to this study, provided the quantitative data used in building analytical models to support the overall theoretical model to be derived from this study.

Observations were also recorded to assess how non-Roman alphabet information search users use such systems with concomitant indications of their preferences for the retrieval and display of bibliographic surrogates. Not under consideration at this time is a potential goal for future research to evaluate the usefulness of commercial machine translation software, which enables users to translate web content using such services as Systran, Google's language tool, WorldLingo, Dictionary.com or other available web tools, including Google's "Translate this page." Evaluating different translation systems is beyond the scope of this study. However, the availability of translation alternatives does enter into this study as an important asset when seeking multilanguage information. Translation is also a key component for the investigation of the broader issues outlined for a research program in the CLIR area.

The observation and extended interviews within the experimental setting are used to clarify user's behaviors and preference when conducting complex searches across languages, cultures, and alphabets when using different IR systems for different tasks.

4.2.2. Pre-test for the experiment

There were several pre-tests in July 2007 before conducting the actual experiment. The pre-test was conducted using a convenience sample which included individuals classified as researchers, librarians and the general public. The language groups were English, Chinese and Korean. A total of seven pre-tests were done and corrections to instruments and procedures were made continuously as the pre-test was processed.

Originally the design called for assigning three different topics to each subject for

that person to search. The three topics were in different areas requiring different approaches to retrieving information. However, this plan was modified during the pre-test since subjects noted that they had little interest in all three topics and that this would influence their seeking of information. When the pre-tests were conducted, it was revealed that the three assigned topics' familiarity scores were somewhat low (around 30% out of 100%). Thus, the design was modified to allow subjects more latitude in selecting their own topic to search and in selecting the IR system for this search. For the other two topics, subjects were assigned those two search topics and assigned the order in which they would search particular IR systems.

The pre-test yielded some additional changes from the original experimental design in its process and contents. For example, the original design included three user groups: researchers, library professionals, and general public users. But the public users were finally dropped since they expressed difficulties in participating in the experiment due to a lack of experience with bibliographic sources in different languages and a lack of competence in reading English materials. In the post task questionnaire, several adjustments were also made to replace technical terms (jargon to the subjects) with more common words. Some redundant questions were also trimmed from this instrument.

It had been considered that it might be appropriate to install transaction search-log software to store all of the subject's history of searching including moves and browsing. But it was deduced that this was not necessary for the current study. The experiment mainly focuses on subjects' query changes with different languages rather than all other changes in the search such as moves, browsing and other behaviors. Hence, the move to a different language was controlled for in the design of the experiment. Data analysis did not require a need to track moves since the experimental design controlled for this and subjects proceeded using an assigned sequence. It is expected however, that future research may want to track moves if subjects were to move freely, at their own will, across languages and systems.

4.2.3. Sample for the experiment

The sample was obtained using a non-probability, convenience sampling method. The sample includes individuals located via a network of colleagues who are researchers and library professionals who lived in central New Jersey and the greater New York metropolitan area. A total of 32 individuals with differing language backgrounds participated in this experiment from July to December 2007. These language groups are Chinese, Japanese, Korean and English. Each language group has a total 8 individuals and these were further divided to include 4 researchers and 4 library professionals. The study was not funded by outside grants or agencies and was totally the responsibility of the researcher.

In this experiment, the limitation to researchers and library professionals was done for two main reasons: first, control of variability and, second, to use individuals who were active users of bibliographic information via IR systems for their research and work. If the searcher population was not restricted, then the study would require a large increase to the sample size since searcher status would become an uncontrolled variable which would split the sample into smaller sub-samples arranged by user groupings. As noted earlier, public users were excluded from this experiment because it was not easy to find appropriate native CJK subjects who can perform this experiment, which required that subjects could also fully understand the task in English. The other reason for modifying the design is that native CJK public users living in the US usually do not use online library systems. These systems are mostly English bound IR systems where individuals search for non-Roman alphabet information. General public users would rather use information systems managed by their own language groups. For example, one of my pre-test participants who are Korean indicated that she usually searches for a Korean websites called "MissyUSA." It is managed by and for the Korean community living in US in order to share information. A Korean who is seeking Korean information might use that country's major search engines such as Naver.com or Korean Google. Thus, these individuals would avoid English systems and this study was designed to investigate how searchers move across languages and IR systems when seeking information. The researcher had recently been a reference librarian in one of New Jersey's busiest public libraries and had observed that Arabic users, Chinese, and even Japanese would often search exclusively in their native language.

Language background	Researchers	Librarian	Total
English	4	4	8
Chinese	4	4	8
Japanese	4	4	8
Korean	4	4	8
Total	16	16	32

Table 3. Subjects' sample distribution by language and field

4.2.4. Experimental design

Unit of analysis is each search. Each subject was assigned three topics and asked to search each in three different systems. Thus, each subject provided data for 9 searches for analysis. Thus, the total N for the study is based on 32 subjects times 9 searches, or N=288. Language rotations were used within each search for comparative purposes.

Subject (Subject $1 \rightarrow 32$)	Topic	System	Language Assigned	Searches	Total searches (unit of analysis)
1	1	1/2/3	C/J/K/E	3	3
	2	2/1/3	C/J/K/E	3	6
	3	3/2/4*	C/J/K/E	3	9
2	2	3/1/2	C/J/K/E	3	12
	1	2/1/3	C/J/K/E	3	15
	3	1/2/4	C/J/K/E	3	18
3	3	4/1/3	C/J/K/E	3	21
	••	••	••		••
	••	••			··
Total 32 subjects					N=288

Table 4. Overall Design of experiment

* System 4: There was one task for each subject where they can choose any IR system for topic 3 which is also subjects' choice of topic. The subject chooses topic and system.

4.2.5. Topics for experiment

Each subject searched three topics: two assigned topics and one topic chosen by subjects in their own area of interest. This last topic is linked to the concept of topic familiarity and it was posited that this might provide interesting results when analyzing data since familiarity is linked to IR search behavior by users. The assigned topics are now described. Note that both topics were created with the following points in mind: (1) the topic should have a research aspect where there would be scholarly data in IR systems; and, (2) the topics have a health aspect and are topics which are also covered in the popular press which makes it likely that subjects would know something about the topic (that is, it would not be too obscure to impose on individuals for IR searching).

1. Topic 1: Fluoride and Health

Please assume that you need information about the effect of fluoride on health. Fluoride is added to water, toothpaste and other products and research indicates it has an effect on health. It is important that reputable and authoritative information be retrieved to address this issue. One report indicated that the best information in this area was contained in (either one of CJK languages will be assigned). Be sure to get high quality information on this topic.

2. Topic 2: Ethnic cook book for health and longevity

Please assume that you need information about Asian health foods, especially using oriental herbs or medicines. It was learned that certain Asian foods are very good at maintaining a person's health and ensuring that the individual lives a very long time. It is important that reputable and authoritative information be retrieved to address this issue. One report indicated that the best information in this area was contained in (either one of CJK languages will be assigned). Be sure to get high quality information on this topic.

3. Topic 3: any topic you have been interested in.

The following table shows actual queries generated by subjects when they search for non-English information.

Category	Queries
Business	Financing Japanese planning
	Marketing computer game in Japan
Library & Information Science	Information retrieval & evaluation
	Music & cataloging & language
	School library service young adults
	Library & google & collaboration
	Korean online bookstore
History	Tokukawaeayasu (Japanese person's name)

Table 5. Subjects' self-generated topics and queries in the experiment

	Daejoyoung (a Korean general)
	Tang dynasty
	Chinese history
Parenting	Traditional parent child relationship
	Children & liar
Culture	Intangible cultural assets
	Cultural properties
	Online shopping behavior in china
Law	Rule of law in china and resident relocation
Technology	Mobile telephones & Asian countries
	Technology & young Asian
Medical/Health/Food	Stem cell
	Smoking & breast cancel
	dyslexia
	Sars
	Spienda
	Tofu & cookery
	Diet
	Food contamination
	Green tea & diet
Culture / Media	Korean soap opera
	Internet alternative media
	Koran popular culture & ty programming
	Japanese drama
	Chinese film
Music	Chinese opera
	History of variation form
	The origin of the piano
Politics	Iraq war
	President of Korea
	Urban politics
	Political relationship in china and Korea
	President lee visit Japan
Fashion/beauty	Lasic surgery & eye
5	Plastic surgery
	Street fashion
	Woman's fashion trend in Korea
Sport	Baseball / ichiro
L	Sumo
Literature	The ending of Harry potter
	Writers in Korea
Tour	Hot spring in Japan
	Sightseeing in China

4.2.6. Systems for the experiment

There is not yet a perfect cross language retrieval system available to public searchers. Although some systems have cross language capabilities, they are not yet perfected to accommodate queries or results in multiple languages. Furthermore, they rarely display results in multiple languages.

In this experiment, three multilingual systems were used: System 1: WorldCat (Online library catalog); System 2: EBSCO Host database (Online database); System 3: Google Language Tools (Commercial search engine). As mentioned in the experimental design, there was one task asked to subjects where they could select any system they wanted to use for their own choice of topic. Seven subjects out of 32 chose different systems than system 1, 2, and 3. Detailed information on this is given later in the data analysis chapter.

Although these systems do not have a perfect cross language retrieval function, they cover non-Roman alphabet information in their collections and have limited language search functions and interfaces. These three systems are very different in their functionality and in the ways in which they display results. These differences affected subjects' searches and use when seeking information using each system.

The WorldCat system is a typical omnibus online public catalog which includes over 95 million records using over 400 different languages. WorldCat provides multilingual access to information in different languages allowing users to search and retrieve in a specified language. Since August 2006, WorldCat has offered free web access to their collection.

EBSCOhost is a provider of a variety of databases covering many different subject

areas. It is mainly used for academic journal article searches providing access to a large set of databases. It provides a multilingual interface and non-English information. It has started to offer limited translation services for bibliographic information and whole articles from one language to another language. These translation services were not offered when this research was conducted during Summer 2007 using the Rutgers University Library system's access to databases.

Google provides access to many types of information and this study focuses on its cross-language capability. One of the functions offered by Google is the translations of a whole web page featuring the summary text page using [Translate this page]. The other tool offered is Google Language Tools which launched in May 2007 and allows users to search in one language and retrieve in a number of specified languages. These services are provided by installing Systran to allow for Machine translation.

4.2.7. Language assigned for this study's experiment

For purposes of this study, each subject was assigned to search using three different languages among the CJKE languages for each of the three topics. These languages are the subject's native language (CJKE), English, and one of the CJK languages which are <u>not</u> her/his native language. For English speakers, they were assigned to search two of the CJK languages for each topic. Thus, the distribution of the total number of different language searches is not even. The following table depicts the language search frequencies and respective percentage for the different languages used in this experiment.

		Frequency	Percent	Cumulative Percent
Valid	Chinese	63	21.9	21.9
	Japanese	67	23.3	45.1
	Korean	69	24.0	69.1
	English	89	30.9	100.0
	Total	288	100.0	

Table 6. Languages used when searching for this experiment

4.2.8. Procedures used in the experiment

The experiment was done at a place convenient for the respondent where there is access to the bibliographic databases selected for this study and where subjects can be observed and interviewed. The place included schools, libraries and cafés where internet service was offered. The experiment was conducted one by one using the subject's choice of computer. The experimental procedure involved the following steps:

- A brief introduction explaining the experiment was provided to the subject. Then, the subject was informed of his/her rights under the protection of human subjects' protocol. The subject was asked to sign the consent form which had been approved by the Rutgers University Institutional Review Board (See appendix 1.1).
- Task instruction sheet was provided to the subject who was asked to review/read it (See appendix 1.2).
- 3) Subject conducted nine search tasks. Each task involved conducting searches for specific information using assigned topic, system and language. While subjects were doing the experiment, the researcher asked subjects to think-aloud and the researcher took notes about their query changes including the language they used,

search time, use of the interface, how many records they read, and any comments from them. Also noted were their responses to any questions posed by the researcher. They had been informed previously that the researcher may ask questions from time to time or ask for an expansion of a think-aloud comment (See appendix 1.3).

- 4) Post-task questionnaire: After each task, the subject filled out a post-task questionnaire. It includes topic knowledge (0-100%), understanding level of bibliographic record (0-100%), intention to use (Yes or No), relevance (0-100%), efficiency searching for non-English information (0-100%), degree of difficulty in conducting the search related to languages, using the interface, judging the surrogates, satisfaction scores for each search, satisfaction about results, satisfaction with each system (See appendix 1.4).
- 5) Post-task interview: After the subject finished all tasks, a brief interview was held in order to get more detailed information about subjects' non-Roman alphabet information search experiences. This interview also asked about their feelings when they encountered any issues or problem constructing queries, judging surrogates, and dealing with language differences (See appendix 1.5).
- 6) Exit questionnaire: After subject finished all nine tasks and interview, subjects were asked to fill out a questionnaire regarding IR system use experience, system knowledge, subjects' background information, such as demographic information, language knowledge, job experience, education level, and age (See appendix 1.6).

4.3. Online survey

This section now reports on the parallel study done: the online questionnaire.

4.3.1. Overview of online survey

This survey was designed to explore non-English information seekers' information needs and study their information seeking behaviors, especially users who use information in non-Roman alphabet languages, such as Chinese, Japanese or Korean. It also assesses if new features would be needed to improve cross-language access to bibliographic records in online catalogs, scholarly databases, and web portals for non-English information seekers. Additionally, it attempted to determine the appropriateness of the bibliographic record retrieved from IR systems.

A total of 204 respondents participated in the online survey with individuals representing academic researchers, library personnel, and the general public. The respondents were a non-probability, convenience sample solicited through a network of colleagues of individuals who had search experience using non-Roman alphabet languages. The survey questionnaire was conducted online using SurveyMonkey through its basic contract by paying a monthly fee to gain access to survey services needed for this study (www.surveymonkey.com). The average participant spent 50 minutes completing his/her survey. It is noteworthy that many participants provided very detailed information in open-ended questions and it can be assumed that the participants had much interest in this topic area.

4.3.2. Pre-test for online survey

In June 2007, 30 pre-test questionnaires were administered using a convenience sample which included researchers, librarians and public users. Many constructive suggestions were received regarding the questionnaire's content, terminology, and unnecessary use of technical language (i.e., clarification of jargon). Several revisions were made during the pre-test phase of the survey study. Librarians at the East Asian Library of Columbia University with a specialty in CJK bibliographic information were especially helpful in providing valuable insights in formulating and revising the questionnaire.

As a result of the pre-test, it was determined that public users needed to be excluded from the administration of the online questionnaire. This was similar to the decision taken with the subjects used in the experiment. One of the reasons for excluding public users from the survey (especially CJK native speakers) was their expressed difficulties in responding to the survey in English. Translating the online survey to CJK languages was considered but finally dropped since it would introduce extraneous variability to the data to be obtained in this part of the research. Instead, the focus would be on individuals who searched for CJK information but who also knew English. The questionnaire was concerned about CJK and English search and retrieval issues. So, researchers and librarians who are the major users of CJK bibliographic information became the target audience for the online survey. For public English speakers, this survey is beyond their experience to answer some of the questions, such as the non-English search experience section in the survey. So, in summary, this survey is targeting specifically those who have experience searching for information written in CJK languages in an English dominant environment.

Various technical problems arose with the SurveyMonkey interface such as displaying captured records from WorldCat. Several contacts with SurveyMonkey resolved these problems.

A preliminary raw data set was prepared using 30 initial responses to assess the viability of the data analysis matrix. These results revealed that some modifications were needed, especially in response categories related to data coding. After modifying the questionnaire to resolve these issues, the final survey was launched from July to September 2007.

4.3.3. Sample and procedures for the online questionnaire

The sample included researchers and librarians who use bibliographic information for Chinese, Japanese, Korean (CJK) and English records. The subjects were those who expected to seek non-English information regularly, those who interacted with non-English databases, and those who conducted research requiring non-English information. Many of the respondents shared all three characteristics of such users.

The subjects were limited to those individuals in the United States and excluded those persons in CJK language-use countries since they often search monolingual IR systems and do not interact as much with English, a major component in the design of this study and a major area covered by the online survey. A prior study by the researcher had explored the use of CJK information by researchers and librarians within the US and in CJK countries (Ha, 2005). In that survey, the subjects were selected who are searching for non-Roman alphabet (especially CJK) information with an emphasis on those seeking information in non-Roman CJK alphabets resident within English environment IR/database systems. Also, the researcher's previous research revealed that CJK language countries have efficient environments to search for CJK language information within a monolingual system (Ha, 2005)

Researchers for the current study were recruited from academic institutions that have East Asian studies, Asian studies, and Asian Languages and Literature programs in the United States. For recruiting researchers, Asian American Net (www.asianamerican.net) was also contacted, which is an organization of Asian studies in the US. This organization posted the survey for their members on their homepage. An email with the survey link was sent to each program or institution's director (or corresponding person) to ask if the researcher could make contact with the faculty members and students in their program. The email was sent about 80 institutions in July and also September 2007. When the initial email was sent, some schools requested that the email be resent after the Fall semester begins. So, by September, all emails were sent again to all schools. About 1/3 of the institutions responded allowing the researcher to contact them or offering that they could post the survey link. A total of 99 researchers participated in this survey (48.5% of the total N=204 respondents). Note that of the 204 respondents, 44 individuals (21.6% of all respondents) did not specify their current position or title.

For the library professionals, the survey was distributed to East Asian libraries with assistance from the CJK working groups in the Council on East Asian Libraries (CEAL) (http://www.eastasianlib.org). Each language group's executive board was contacted and asked to distribute the online survey. There are more than 80 East Asian libraries in the

United States. This survey included a total of 64 librarians (31.4% of all respondents).

4.3.4. Questionnaire details

The questionnaire contained four sections as follows (See appendix 2).

- Online searching experience: participants' overall online search experience and their information use experiences with OPACs, WorldCat, DBs, search engines, and translation systems.
- 2) Non-English information searching: detailed questions about participants' experience with non-English searching using various systems. Non-English information needs, expectations and difficulties in seeking such information were also incorporated into this section.
- 3) Bibliographic record evaluation: asked to evaluate usability of different types of bibliographic records, especially its language representation. Four different types of records were evaluated. For example, one bibliographic record shows only Romanized description, whereas the other shows vernacular (original language) description with English translation in addition to Romanized bibliographic record. The sample bibliographic records were randomly selected from WorldCat. Participants provided an evaluation of their understanding level for the bibliographic record on a 7 point scale. Respondents also provided a score to indicate their intention to use the record to secure the document described.
- Background questions: participants' background information to include language background, language knowledge, education level, current job, and

area of specialty. It also asked if they knew non-English journal names.

4.3.5. Survey participants' background information

Total survey participants included 204 individuals although some people did not reply to all of the questions. The table below gives a summary of response rates for the background information component of the online questionnaire.

Category	Descriptive statistics
Total survey participants: N=204	
Native language (N=170)	English (n=84) 49 %
	Korean (n=33) 19 %
	Chinese (n=31) 18 %
	Japanese (n=12) 7%
	Others (n=10) 6%
Current employment (N=167)	$P_{asaarabars}$ (n=101) 60%
Current employment (N=107)	student $(n-61)$ 28%
	doctoral student $(n-45)$ 27%
	- doctor at students $(n-7)$ 49/
	- master students $(n-1)$ 4%
	- conege student (n=10) 0%
	researcher $(n-11)$ 6%
	researcher (n=11) 0/0
	Library professional (n=66) 40%
	Librarian (n=50) 30%
	Director/manages in information
	center (n=10) 6%
	Library assistant (n=7) 4%
Very effect energine of (N_160)	M 12.57
rears of job experience (N=160)	M=12.57
The number of languages known	M=3.94, Mode=4
(including mother language) (N=161)	(85% of survey participants know
	2-5 languages)
	The mode is 4 languages: (n=45) 28%
Education level ($N=165$)	Ph D $(n=41)$ 25%
	1 11.12. (11-71) 20/0

Table 7. Summary of survey participants' background information

	Master (n=99) 60% BA (n=25) 15%
Years of staying in USA (for those whose nationality is not US) (N=85)	M=9.35 Mode=5
Age (N=162)	M=41 yeas old, Mode=35 years old
Gender (N=163)	Female: (96) 58.9% Male: (67) 41.1%

4.4. Data analysis methods

Table 8. Data analysis plan by each research question & hypothesis

Research questions & Hypotheses	Method
1 What are non-Roman alphabet users'	R1.1. Quantitative (descriptive statistics) and categorical groupings
information system use and search	(categorizations) / survey questionnaire. Basic descriptive statistics and summary
behaviors?	of participants' comments will be presented from survey results.
1.1 How do users use current information	
retrieval systems searching for non-Roman	R1.2. Categorical groupings (categorizations) / survey questionnaire
alphabet information?	Based on survey participants' comments, major concepts will be extracted and
1.2 What are the users' needs and	presented.
expectations when accessing non-Roman	
alphabet information using such IR systems?	H1.3. Quantitative /survey & experiment (descriptive statistics and oneway
1.3 What are their language choices and	analysis of variance by subjects' language background)
considerations when they construct their	- Survey: preferred language (English/native language/target language) in
queries?	1. English environment, 2. non-English environment
H1.3: Users will input (or prefer) to	- Experiment: preferred language (English/native language/native & target
state queries in their native language rather	language/target language)
than English.	
1.4 How do users' background and their	R1.4. Qualitative and quantitative analysis / experiment
experiential knowledge affect their non-	By observing participants' use of systems, unique behaviors by users'
Roman alphabet information system use and	background will be explored.
search behavior, if at all?	
H1.4: There are statistically significant	H1.4. Quantitative analysis / experiment
differences in users' IR system use and	ANOVAs (oneway Anova with multiple group comparisons) and Chi-Square for
search behavior by users' background and	nominal variables
their experiential knowledge.	- IVs (Factor): 1. Users' background (language background, career, education)
	2. Experiential knowledge (language knowledge, search
	experience, topic knowledge)
	- DVs: Search behavior (number of terms used, number of query changed, search

 1.5 Are their user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine? H1.5: There are statistically significant differences in users' IR system use and search behavior for non-English information when compared to English language searching. 	 time, Boolean logic use, language choice, and advanced tool used) H1.5.1 Quantitative / experiment by query analysis One way ANOVA by English vs. non-English searching - IVs (Factor): non-English search vs. English search - DVs: search behavior (number of terms used, number of query changed, search time, Boolean logic use, language choice, and advanced tool used) and system evaluation.
 What issues are present when individuals search non-English information via online DBs, OPACs and web search engine? What kinds of issues and limitations exist when searching non-English information? How do those factors affect the access to information contained in such systems? H2.2.1: Users will avoid databases if the system lacks efficient and comprehensive language coverage when they look for non-English information. H2.2.2: If the system lacks efficient and comprehensive language coverage, then users' search time will be shorter than with other systems when they search for non-English information. H2.2.3: Users have better knowledge about scholarly academic journals in their areas written in English than journals written in non-English. 	 R 2.1: Qualitative descriptions from survey results H 2.2.1: Quantitative / experiment Chi-Square test in users' choice of IR system An ANOVA test in different search behaviors by three different systems H 2.2.2: Quantitative / experiment An ANOVA test in search time by three systems A logistic regression model will be tested to see which variables are related to explain a variable "gives up" (binominal) of a search (i.e, terminates the search), and how much variability is explained in subjects' termination of the search. H 2.2.3: Qualitative description from survey results

2 Harry de biblier mentie manual facilitate en	112.1. Organitation 9 and literious from any mult
3. How do bibliographic records facilitate or	H3.1: Quantitative & qualitative from survey result
minder the understanding of bibliographic and	Descriptive statistics and summary of survey participants comments
II2.1. There will be statistically	112.2. Orangi institution and train from a second instant
H3.1: There will be statistically	H3.2: Quantitative analysis from experiment
significant differences for users' understanding	One way ANOVA by English vs. non-English searching
levels searching with full record which includes	
English translation, Romanized and vernacular	H3.3: Quantitative analysis from survey
language than with other record displays.	Multivariate analysis of variance (MANOVA, or multivariate GLM) in 4 types of
H3.2: There will be statistically	bibliographic record understanding levels by researchers vs. library professionals.
significant differences in bibliographic	
understanding level by English searching	
compared to non-English searching.	
H3.3: There will be statistically	
significant differences in bibliographic	
understanding level by researchers compared to	
library professionals.	
4. What explains non-English information	R4: Quantitative /experiment
users' relevance certainty judgments (intention	A logistic regression model will be tested to see which variables are related to
to use retrieved information)? How do these	explain the decision making (intention to use after reviewing a bibliographic
intentions compare to those reported for English	record), and how much variability in that decision whether to use the retrieved
or monolingual users?	record or not.
	The logistic regression will yield the same prediction table as a discriminate
	analysis and this will indicate the probability of prediction whether users will
	choose a specific bibliographic record or not
	choose a specific biolographic record of not.
	Regression, t-tests, and ANOVA will be used as exploratory tools to investigate

H4 Caveat: User characteristics may account for some of the variance when making relevance judgments.	H4 Caveat: User characteristics may account for some of the variance when making relevance judgments.
H4: There is a statistically significant difference in relevance certainty judgments for searches seeking non-English information compared to searches seeking English information.	H4: Quantitative /experiment Oneway ANOVA test on different language searches
5. What would constitute an explanatory model for users searching for non-English information? How does this model explain influences of user characteristics, language, topic, and task in their query formulation and relevance evaluations of bibliographic information?	This calls for revising and re-casting the original theoretical model to reflect the empirical relationships established by hypothesis testing. The recast model can then be used in future studies.

Chapter 5: Data Analysis Results

In this chapter, the results of both the experiment and online survey data analyses will be reported. These results are linked to their corresponding research questions and hypotheses provided in earlier chapters.

5.1. Research question 1

1 What are the patterns of non-Roman alphabet users' information seeking and use behaviors?

- 1.1 How do users use current information retrieval systems when searching for non-Roman alphabet information, especially via online DBs, OPACs and web search engines?
- 1.2 What are the users' needs and expectations when accessing non-Roman alphabet information using such IR systems?
- 1.3 What are their language choices and considerations when they construct their queries?
- 1.4 How do users' language and background affect their non-Roman alphabet information seeking and use, if at all?
- 1.5 Are their user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine?

R1.1. How do users use current information retrieval systems searching for non-Roman alphabet information, especially via online DBs, OPACs and web search engines?

Online survey participants (N=192) provided their current information system use. The participants' average online searching experience is 12 years (M = 11.84, SD = 5.937). Most respondents (102 people, 53%) answered they have bibliographic record search experience as recently as one day ago or today. Surprisingly, 40% (n=76) of participants of this survey answered that they search non-Roman alphabet information everyday. About 56% (n=107) searched non-Roman information within the last week. Only 16% (n=31) answered that they have never searched non-English information.

Overall, participants in this study are very experienced in using online searching and many use bibliographic information on a daily base. Importantly, with the exception of only 31 participants (16%), most survey respondents (86%, n=134 out of 192) have experience with non-Roman alphabet information searching within the last month. As already noted, among this subgroup of individuals 40% (n=76) of the participants search non-Roman alphabet information on a daily base.

Major survey participants' system use results are reported as follows.

1.1.1 .Survey participants' online sources experience for their general searching and non-English information searching was recorded using the 7 point scale.

	For general searching (N=192)			For non-English searching (N=175)		
	М	SD	Mode	М	SD	Mode
Library catalog	5.80	1.74	7 (n=105, 55%)	4.03	2.25	6 (n=43, 24%)
WorldCat/RLIN	3.78	2.43	1 (n=58, 30%)	3.41	2.37	1 (n=65, 37%)
Online Journal	4.94	2.05	7 (n=66, 34%)	3.28	2.19	1 (n=57, 33%)
Web Search engine	6.55	.98	7 (n=145, 76%)	5.58	1.73	7 (n=72, 42%)

The survey participants use library catalogs on an almost daily basis and most

Table 9. Survey participants' experience of online sources (not at all = 1/most use = 7)

respondents answered they use the library near their home location which can include public libraries. Experience with WorldCat or RLIN system shows an almost bimodal pattern with those systems used either 'not at all (30%)' or 'most use' and 'very often use' (34%). It is assumed that the WorldCat or RLIN system is well known for library professionals but not as well known for the researchers. This was also reconfirmed from the experiment subjects' experience (16 researchers and 16 library professionals). Only 3 researchers among 16 researchers had experience with WorldCat and RLIN systems. Use of online databases to access journal information shows that most participants (61%) report that they use such IR systems very often. Also revealed was the widespread use of web search engines by survey participants. The most used web search engine is 'Google' (86%).

For non-English information searching (N=175), only 24% (n=43) of the participants responded they are using library catalogs very often and 21% (n=38 out of 175) of the participants answered they have never had OPAC experience searching for non-English information. While half the participants mentioned they use their local library, the rest revealed that they use particular OPAC systems which are nationally known for their Asian collections, such as the library catalog systems at Columbia

University, Princeton University, Cornell University, Harvard University, UC library system (MELVL), and the OCLC system. Some participants even mentioned library systems which are located in other countries, such as Yonsei University library system and KERIS in South Korea, Washeda University and National Diet Library (NDL) in Japan, and several other libraries.

The sample was divided in its use and familiarity with WorldCat or RLIN with many individuals (37%, n=65) answering they have no experience with WorldCat or RLIN, whereas another group of respondents (34.6%, n=61) report extensive experience with these systems. The majority of respondents, 57.6% (99 out of 175), reported that they do not seek Non-English journal articles using database systems. The mean response on this 7-point scale was low (M=3.28, SD=2.193, Mode=1). This is very different from these individuals' general information searching (M=5, SD=2, Mode=7). Only 23% of the participants responded they use online DBs very often to seek journal information. For those who sought information for non-English journals, 80 respondents identified specific DBs and many of these were repeated by the survey group: "China Academic Journal" and "JSTOR" which were mentioned more than 10 times and "Magazine Plus," "CiNii,"

As can be seen in table 8, web search engines were used the most when individuals sought non-English information. Google is dominant among such sources. Several respondents indicated that they search regional Google interfaces, such as "google.tw.com" which is Google for Taiwan. "Baidu" was mentioned from several respondents which provides multimedia information and is searchable in vernacular languages including CJK. Several regional search engines were mentioned too, such as "Naver.com," and "Daum.net" from South Korea.

From this result, it can be seen that the clearest difference between general searching and non-English searching is users' online journal searching experience. Note the sharp comparison for general searching (mode = 7 indicating most use) with non-English journal searching (mode = 1 indicating no use). It might be assumed that either they are not looking for journal articles written in non-English languages or they do not know which online database accesses such information. It is also possible that they assume that there is a paucity of online databases covering non-English materials. Later, this issue will be addressed again.

1.1.2. Experience in searching for information in languages not known to the respondent.

The length of average time since searching for such information was 71 days ago (M = 71.09, SD = 216.566, N=192). Many survey participants (n= 69, 35.9%) indicated that they have never sought information in languages unknown to them. Yet, others do conduct such searches. Within the last week, there were 50 (26%) subjects who searched for information in languages not known to them. Across the sample, individuals conducted numerous searches in different languages. More than 30 languages were searched and among them Chinese, Japanese, and Korean languages were the most frequently used (46% of respondents searched in these languages). This was followed by German and Russian languages (10%).

For those searching in languages not known to them, the purpose of the search

included "research" (40%, n= 48) and "library work" (31%, n=39). Some other reasons were:

a. found web results or related articles that included a not known language:

"When I searched the topic in English, a retrieved document was written in German"; "only available in that language"

"only available in that language"

b. "I wouldn't use translation systems. I can't trust current translation system."

1.1.3. Search format for non-English information (1: not at all – 7: most use)

The following table 10 gives the search focus for those seeking non-English materials. Note that the highest focus of non-English seekers was for internet resources (M=5.48).

Search focus	n	Mean	S. D.
non-English academic journal	175	4.27	2.317
non-English news/magazine article	176	4.49	2.017
non-English book	176	4.91	2.141
non-English internet resource	176	5.48	1.603
non-English media	168	3.76	2.134

1.1.4. System use for each different format

There were 148 individuals who responded to what kinds of online resources they use for a specific format when searching for non-English information. As shown in the following table 11 and as expected, library catalogs and the worldwide union catalog was most often used to find books. As noted earlier, respondents use web search engines more than any other system when searching for non-English information on the Internet.

	Academic	News/magazine	Books	Internet	Media (movie,	count		
	journal	articles		resources	music etc)			
	articles							
OPAC	32% (48)	15% (22)	70%	13% (19)	24% (36)	227		
			(102)					
Worldwide	18% (27)	5% (8)	60%	9.4% (14)	17% (26)	164		
union catalog			(89)					
Online DB	70% (102)	32% (48)	7%	4.7% (7)	2.7% (4)	172		
journals			(11)					
Web search	38.5% (57)	57% (85)	42%	90.5%	55% (82)	420		
engine			(62)	(134)	. ,			

Table 11. System use for different formats when searching for non-English information

1.1.5 A question was asked to identify what is the most often used system when searching for non-English information (N=160).

There were 85 survey participants (53%) who responded that Google constituted the most used web search engine when seeking non-English information. The OCLC system with WorldCat followed next but with a much lower use patterns. Only 9.4% (n=15) answered they use the OCLC system the most when they need to search for non-English information. More than 50 different systems were mentioned and table 12 below gives a summary of this finding.

The reasons for using Google were for its convenience in accessing non-English information. Several individuals mentioned that Google is a good starting tool since it gives an indication of what might be available on a topic in a non-English language. Respondents indicated that other systems were used when the user was familiar with the system's language functions. Participants whose native languages are not English use their native search engines because they are familiar with them and feel that it yields more reliable information than other systems. The following table 11 shows survey
participants' comments about each system indicating why users prefer to use a specific

system when searching for non-English information.

Table 12. Reasons given for respondent's most used systems when seeking non-English information

System	Use reason
Google	 Convenience, ease of use, easy access, the quickest
	Good for initial stages of search: "As it covers more
	exhaustive info, I use Google first and then move on to another
	system to specify the info I need."
	No choice: lack of better systems
	Language availability:
	"It's the only search system that I know having provided
	multi-language interface."
	"I am able to search in any language I need."
	"It allows me to type in non-Latin script. I can also change
	the interface if needed to that same script. I can specify only
	items from a particular country. The results are shown in the
	language of origin."
	"You have a "Click-on option" to translate the web page."
	 Variety of format/subjects: Includes the most formats for
	general searches in any subject area.
	• Global access: Easy way to access information organized by
	and intended for native speakers of that language (ie. Chinese
	websites).
OCLC (WorldCat)	• The most comprehensive source: "larger database than local /
	biggest union catalog in the world; we use world at to help
	of non English materials is near "
	I anguage function: "because it's bilingual: most likely to have
	data in vernacular AND romanized forms ": "I am usually
	searching for books in Japanese, and Lean use Japanese
	characters in WorldCat to find them work related "
RI IN Eureka	"good language function: presents the non-English on same
REIT V EUTOKU	screen as transliterated data"
Amazon.co.ip	"It has a huge selection of books and includes descriptions and
J	reviews."
NDL-OPAC	"comprehensive search, interface is easy to master, often
	provides readings of Japanese names
China Academic	"Full-text of Chinese articles"
Journals	"English input and give me Chinese text"
CNKI	"dedicated database of Chinese language journal articles"

Wikipedia	"It provides a fairly accurate way to locate non-English
	information very quickly through its built-in interwiki language
	links."
	"Best use of UTF-8 & web standards; most reliable access."
NACSIS	"provides information on almost everything published in
	Japanese."
Webcat Plus	"Because I can search in Japanese"
Ritsumeikan	"I start here because I am familiar with the system; also provides
University OPAC	phonetic spelling in katakana for all kanji in article and book
	titles"
UCLA library	"very well synchronized for CJK language sources"
catalog	
PubMed/Medline	"Covers 85% of global biomedical, and healthcare related peer-
	reviewed literature"
Naver	"This system is more familiar than other online systems because
	this is based on Korea"
	"to find information saved in Korean"
	"Has information that I can't get on other systems"
KERIS	"comprehensive; name is familiar; accessible"

1.1.6. Any CLIR function system that they know?

Few respondents indicated that they knew that Google has language search tools which allow it to function within a CLIR mode. Some survey participants were aware that China Academic Journals (CAJ) allows the searcher to type queries in English and return full-texts in Chinese. CAJ also provides English abstracts.

Meanwhile, many respondents seem to be confused about what constitutes a CLIR function since they mentioned WorldCat, Yahoo, Baidu, wikipidia as having a CLIR function. Baidu is the only one among these web search engines tools that can read and retrieve pinyin words, that is, Chinese Romanization, as WorldCat does for Chinese. Several respondents were convinced that certain academic library systems (such as UCLA, Duke, University of Chicago) were capable of providing a CLIR function. Note that these systems do allow limited vernacular retrieval functions since they have well organized bibliographic records, vernacular and Romanized access, and English records. Within such systems, users can search by vernacular language, Romanized words, or English if that word is designated in the record. Finally it should be noted at this point that there are few fully functional CLIR systems now available.

R 1.2 What are the users' needs and expectations when accessing non-English information when using such systems?

1.2.1. Users' need for non-English information

Survey respondents provided information regarding circumstances when they might need to access or use information written in non-English languages. Selected representative comments are summarized as follows:

Category	Comment
Research	"related to my research, there might be good source written
	in other languages."
	"In order to expand the list of the literature that I can
	utilize"
	"when doing cross cultural searches"
	"to increase understanding for my research, I search
	information in my native language"
	"for my foreign language class"
Lost translation	"because some of the message in the original language
- can't trust translation	cannot be translated into other languages and therefore
- want to see real work	becomes a loss of value. It is worthwhile to go back to the
	source language and try to understand the meanings of the
	work that is true to the author's intent."

Table 13. Users' need for non-English information

	"when trying to verify the accuracy of information (factual
	or interpretive) presented in a translated text."
The only one	"when it is the only source of information or when the
	information in my own language is not sufficient, which is
	often the case"
Work	- For library work: to assist library patron / library catalog
	& collection
	- For business (for work)
Curiosity and respect to	"the information written in languages I am not familiar
other languages	with is as important as the one in familiar languages
	because it might be crucial to someone"
	"material from different language version might carry
	additional/different content"
	"to look from the different point of views and supplement
	each other."
	"when seeking different interpretations and perspectives
	other than English-speaking countries. E.g., reading stuff
	about 9/11 and Iraq War from the other countries'
	perspectives"
To get information about	"I am not a native English speaker. I want to access news,
my home country	especially for my home country news, in my language"
For personal interest	Including information for fun
	Japanese video game information
For general information	Health issues etc Oriental medicine in Asian

There were 97 subjects (out of 192, 50.5%) who responded that they need non-English information for their research. "Personal interests" was also included in the need of non-English information by 32 subjects (16.7%). One subject mentioned there is always advanced information about video games from a certain country. Respondents noted that library work increasingly serves diverse patrons with its collection and at the reference desk.

1.2.2. Non-English information users' expectations

Survey participants were also asked what they expect when using IR systems to search for non-English information. They mentioned such topics as the system's retrieval

function and its interface design, but all expectations can be summarized in two areas:

1) one is they want to have a system that can search and retrieve across languages;

2) the other is they want to have informative bibliographic records which can be

understandable to people who do not know the target language.

For these respondents, who knew English, the understandable description could be

presented with English translations including an English abstract. The number of

worldwide users of IR systems who can read English might be inferred from languages

used in the World Wide Web.⁵

⁵ It is estimated that 68% of World Wide Web content is in English. The next most common language is Japanese representing 5.9% of web content, followed by German with 5.8% of web content, and then Chinese at 3.9%. Source: <u>http://global-reach.biz/globstats/refs.php3</u> (accessed March 19, 2008). See pie diagram depicting languages' content on the World Wide Web.



Web Page Title: Global Reach. Global Internet Statistics (by Language). [Note that this chart represents the language of individuals using the Web regardless of the country they are in. The chart's slice to the left of English represents combined Other languages.]

Source: http://www.glreach.com/globstats/index.php3 (accessed March 19, 2008).

Category	Comments
Cross Language	"I hope that it won't matter what language I use. It will
Information Retrieval	just show up in my search result"
function needed	"Automatic cross-language links is definitely a necessity
	although I know it is still very much flawed"
	"The ability to search in English for other languages on
	search engines (i.e., translate your query to that
	language)."
Record display	- original and English translation transcription
	- English abstract/summary
	- Table of Content (TOC)
Easy access to full text	- availability of database covering non-English materials
non-English information	- full text
Translation service	"Add a service with free translation service."
assistance	"Convenient online dictionary"
	"I need PERFECT translation"
Facilitate multilingual	- Multiple languages support
query input/search	- Query input help
methods.	- standard Romanization access/
	- help
	- nuanced description by working at controlled
	vocabulary

Table 14. Non-English users' expectation for system

R 1.3. What are their language choices and considerations when they construct their queries?

H1.3: Users will input (or prefer) to state queries in their native language

rather than English.

A. From experiment results

When they are asked to search non-English information (CJK), most subjects' first

language input choice was English (90%). It can be assumed it is because in this

particular experiment, subjects were mostly students and librarians who know English

and who are living in the US. Also, the experiment was primarily done using an English language environment (i.e. computer process environment and task instruction). Only about 10% of tasks were tried in either 'their native and target language' or their 'native' *or* 'target language.' So, English was the preferred "first choice" language and not the user's native language; thus, this hypothesis was not supported.

Table 15. This hanguage choice from experiment				
Languages	Frequency	Percent	Cumulative Percent	
English	225	89.3	89.3	
native & target language	16	6.3	95.6	
native language	7	2.8	98.4	
target language	4	1.6	100.0	
Total	252	100.0		

Table 15. First language choice from experiment

B. From online survey results

The online survey assessed the preferred language when users seek non-English information using two system environments: English system and non-English system environment. The environment means the system supports English or one of the non-English languages. English was chosen most frequently at a 55% preference (n=98) and this was followed by the target language at 33.7% (n=60). The person's native language preference was reported at 10.7% (n=19) in the English system environment (N=178). So, again but using a different study, the hypothesis is not supported.

Table 16. A prefer search language for non English information in English system (N=178)

Rank	Language	Percent (number of people)
1	English	55.1% (98)
2	Target language	33.7% (60)
3	Native language	10.7% (19)

In the non-English system environment (N=177), respondents chose the 'target language' most frequently at 52% (n=92). Following this was the person's native language at 23.7% (n=42) and then English at 22% (n=39). It is interesting here that the participants chose the target language as the most preferred language. It does makes sense, however, if a system can support advanced language function, as a CLIR function, then the target language can be the first choice of language no matter what languages are available since this is the language where the information is being sought. Thus, the hypothesis is not supported.

RankLanguagePercent (number of people)1Target language52% (92)2Native language23.7% (42)3English22% (39)			
1 Target language 52% (92) 2 Native language 23.7% (42) 3 English 22% (39)	Rank	Language	Percent (number of people)
2 Native language 23.7% (42)	1	Target language	52% (92)
3 English 22% (30)	2	Native language	23.7% (42)
	3	English	22% (39)

Table 17. A prefer search language for non English information in non-English system (N=178).

R 1.4. How do users' background and their experience knowledge affect their non-English information seeking and use, if at all?

H1.4. There are statistically significant differences in users' IR system use and search behavior by the users' background and experiential knowledge.

The hypothesis is supported with many analyses reported below.

1.4.1. Search behavior from experiment observation

Observation and interview data are categorized using data exploration methods where patterns of searching are reported for the respondents. Noted here are differences in search behavior by native CJK subjects and native English speakers.

A. Native CJK subjects' have a better understating of Chinese characters and try them in their queries. Note that this is true of all CJK subjects. They try to input Chinese character as a query if they are assigned any one of the CJK alternatives as their target language. One subject mentioned s/he does not understand "pinying (Chinese pronunciation)" when s/he review the bibliographic record and added, "Since some Koreans can understand Chinese characters, they might be more helpful than pinying" (Subject 4). Japanese use more Chinese characters than Korean do so this applies to Japanese too. It is important to point out that this assists those who know Chinese characters even though the three CJK languages can use different Chinese characters.

B. The native CJK users use their own reference tool to find a right target word(i.e. online dictionary or their native language's web search engines, or by using Google).Many CJK experiment participants also tried those systems to obtain additional

information prior to starting their search.

C. In general, English speakers' attempted to use English to retrieve non-English records without any other action such as looking for options in the interface to find translation functions or referring to other tools. Many native English speakers including librarians were not able to find information written in assigned languages. They would need assistance to access non-English information

1.4.2. Non-English information search behavior by language background in the

experiment

The following table 18 presents descriptive statistics on non-English information searching. Of the total 288 searches, 199 were conducted for non-English information (and this was accomplished partly by the design of the experiment and the selection of subjects).

	Ν	Min	Max	Mean	S. D
The number of different language used	199	0	3	.41	.876
The average number of search terms in one query	199	1	5	2.45	.952
The number of query changed	199	0	17	2.93	3.114
Search time (minutes)	199	0	24	5.69	4.308
Boolean term use between queries	199	0	1	.60	.491
Advanced search option use including limit options	199	0	1	.74	.440
Valid N (listwise)	199				

Table 18. Descriptive Statistics for non-English searching

1.4.2.1. Test for continuous search variables

A. descriptive statistics for numeric and continuous search behaviors

As seen in the table19, there are notable mean differences between subjects with a CJK language background compared to English native subjects. Native English subjects rarely tried different languages in their query formulation (variable 1) and they did not expend much effort searching for non-English information. Other variables (2, 3, and 4) also reflect this same pattern and report a lower mean score for native English subjects compared to CJK language background groups.

	native language	Mean	S. D.	Ν
1. The number of	Chinese	.40	.873	55
different languages	Japanese	.63	1.019	51
(Including Romanized) used	Korean	.61	1.039	44
Romanizeu) useu	English	.00	.001	49
	Total	.41	.876	199
2. The average	Chinese	2.76	.942	55
number of search	Japanese	2.57	.922	51
terms in one query	Korean	2.57	.818	44
TOTTICIALION	English	1.88	.881	49
	Total	2.45	.952	199
3. The number of	Chinese	3.71	3.065	55
query changes	Japanese	2.33	2.636	51
	Korean	4.00	4.092	44
	English	1.73	1.901	49
	Total	2.93	3.114	199
4. Search time (minutes)	Chinese	5.78	3.287	55
	Japanese	5.84	4.397	51
	Korean	7.05	6.019	44
	English	4.20	2.799	49
	Total	5.69	4.308	199

 Table 19. Descriptive Statistics of search behaviors by native language

B. One way Multivariate Analysis of Variance (MANOVA)

One way Multivariate Analysis of Variance (MANOVA) test was conducted to see if there are mean differences in participants' search behaviors by their native language background. Four numeric search variables were tested: 1. The number of different languages (including Romanized) used; 2. The average number of search terms in one query formulation; 3. The number of query changes; and, 4. search time in minutes. The results support the hypothesis that subjects' search behaviors are statistically different by different language background. Note that for this analysis, the search behaviors are represented by the four numeric variables.

Since the Box's test is significant as F(46, 72328) = 2.23, p < .001, the homogeneity hypothesis can be rejected, and it can be assumed that there are differences in the matrices. Also, as seen in the below multivariate table, Wilks' Lambda .754 is significant as F(12, 508) = 4.77, p < .001 and this indicates that the population means for the dependent variables are not the same as for the four different language backgrounds.

Table 20. MANOVA test for search behavior by language background

Between-Subjects Factors of MANOVA for search behavior by language background

		Value Label	Ν
native language	1	Chinese	55
	2	Japanese	51
	3	Korean	44
	4	English	49

Box's Test of Equality of Covariance Matrices (a) of MANOVA

Box's M	46.416
F	2.229
df1	20
df2	72328.700
Sig.	.001

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a Design: Intercept+nativelang

Effect	-	Value	F	Hyp df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.894	404.359(a)	4.000	192.000	.000	.894
	Wilks' Lambda	.106	404.359(a)	4.000	192.000	.000	.894
	Hotelling's Trace	8.424	404.359(a)	4.000	192.000	.000	.894
	Roy's Largest Root	8.424	404.359(a)	4.000	192.000	.000	.894
Native language	Pillai's Trace	.262	4.641	12.000	582.000	.000	.087
33.	Wilks' Lambda	.754	4.767	12.000	508.276	.000	.090
	Hotelling's Trace	.305	4.844	12.000	572.000	.000	.092
	Roy's Largest Root	.216	10.473(b)	4.000	194.000	.000	.178

Multivariate Tests(c) for search behavior by language background

a Exact statistic

b The statistic is an upper bound on F that yields a lower bound on the significance level.

c Design: Intercept+nativelang

Since four variables were tested, each ANOVA was conducted at a significance level of p = 0.012 (.05 divided by the number of ANOVAs conducted). All four variables were significant. Thus, there are statistically significant mean differences in participants' search behaviors by their native language background for languages used, average number of search terms, number of query changes, and search time. Partitioning the variance explained would not be productive at this point since other variables are influencing the model. Instead, later tests will be used to incorporate more variables into the explanatory model. The purpose at this point is to determine if native language affects search behavior. It does, and it does so significantly:

The number of different language used: F(3, 195) = 5.81, p < .001

The average number of search terms: F(3, 195) = 9.47, p < .001

The number of query changes: F(3, 195) = 6.39, p < .001

Search time: *F* (3, 195) = 3.56, *p*<.05

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Native language	The number of different language (including Romanized) used a	12.477	3	4.159	5.811	.001	.082
	The average number of search terms in one query formulation b	22.799	3	7.600	9.469	.000	.127
	The number of query changed c	171.921	3	57.307	6.392	.000	.090
	Search time (minutes) d	190.688	3	63.563	3.558	.015	.052
Error	The number of different languages (including Romanized) used	139.553	195	.716			
	The average number of search terms in one query formulation	156.498	195	.803			
	The number of query changes	1748.230	195	8.965			
	Search time (minutes)	3483.995	195	17.867			

Table 21. Tests of Between-Subjects Effects for search behavior by language background

a R Squared = .082 (Adjusted R Squared = .068)

b R Squared = .127 (Adjusted R Squared = .114)

c R Squared = .090 (Adjusted R Squared = .076)

d R Squared = .052 (Adjusted R Squared = .037)

Post hoc tests were done using the Dunnett method since equal variances were not assumed for each language group given the sample and its characteristics (with each having different sample sizes, (Chinese 55/Japanese 51/Korean 44/English 49). This post hoc results show exactly how and which language groups have statistically different results. Mostly, the differences occur between English background subjects and any one of the CJK subjects. The hypothesis is supported: there are statistically significant differences in users' IR system use and search behavior by users' background.

Dependent Variable	(I) native language	(J) native language	Mean Difference (I-J)	Std. Error	Sig.	98.8% Co Inte	onfidence rval
						Lower Bound	Upper Bound
The number of	Chinese	English	.40	.166	.045	08	.88
different language	Japanese	English	.63(*)	.169	.001	.14	1.12
used	Korean	English	.61(*)	.176	.002	.11	1.12
The average number of search terms in one query formulation	Chinese	English	.89(*)	.176	.000	.38	1.39
	Japanese	English	.69(*)	.179	.000	.17	1.21
	Korean	English	.69(*)	.186	.001	.15	1.23
The number of query	Chinese	English	1.97(*)	.588	.003	.28	3.67
onangoa	Japanese	English	.60	.599	.625	-1.13	2.33
	Korean	English	2.27(*)	.622	.001	.47	4.06
Search time (minutes)	Chinese	English	1.58	.830	.145	82	3.98
	Japanese	English	1.64	.846	.134	80	4.08
	Korean	English	2.84(*)	.878	.004	.31	5.38

Table 22. Post hoc tests for search behavior by each language group (Dunnett t (2-sided))

Based on observed means.

* The mean difference is significant at the .012 level.

a Dunnett t-tests treat one group as a control, and compare all other groups against it.

1.4.2.2. Test for nominal search variables

A. Chi-Square test for Boolean logic use by different language background.

It is noteworthy that 60% of the subjects used Boolean logic when searching for non-English information. Compared to other language groups, native English subjects used less Boolean logic than non-native English subjects (Table 23). This is depicted in a bar chart showing these differences (Figure 3). As previous studies indicated, most subjects use the 'AND' Boolean operator when combining search terms (Marchionini, 1988; Sewell & Teitelbaum, 1986; Trzebiatowski, 1984). When observing experiment participants' search behavior, 60% of them used the 'AND' Boolean operator in the three task systems (WorldCat, EBSCOHost, and Google) in one query box. It almost looked as if some had developed a habit to insert 'AND.' When WorldCat and EBSCOHost did not accept the query with 'AND' in one query box, participants seemed distressed that they should have to insert it one by one in different query boxes on an advanced search screen.

Table 23. Chi-Square test of Boolean term use between queries in non-English searching

		Frequency	Valid Percent
Valid	No	79	39.7
	Yes	120	60.3
	Total	199	100.0

Crosstabulation

				native language			
			Chinese	Japanese	Korean	English	
Boolean	No	Count	20	17	11	31	79
Term use		Expected Count	21.8	20.2	17.5	19.5	79.0
		% within Boolean term use between queries	25.3%	21.5%	13.9%	39.2%	100.0%
	Yes	Count	35	34	33	18	120
		Expected Count	33.2	30.8	26.5	29.5	120.0
		% within Boolean term use between queries	29.2%	28.3%	27.5%	15.0%	100.0%
Total		Count	55	51	44	49	199
		Expected Count	55.0	51.0	44.0	49.0	199.0
		% within Boolean term use between queries	27.6%	25.6%	22.1%	24.6%	100.0%

Chi-Square Tests for Boolean logic use by language background

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.458(a)	3	.001
Likelihood Ratio	16.414	3	.001
Linear-by-Linear Association	5.513	1	.019
N of Valid Cases	199		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 17.47.

The Chi-Square results show that Boolean logic was used differently by different language groups (Pearson Chi-Square (3, N=199) = 16.46, p < .001). So the hypothesis is supported: there is statistically significant difference in users' search behaviors in employing Boolean logic operators by users' background.



Figure 3. Boolean logic uses by language background

B. Advanced search option use by different language background.

Advanced search option was used in 74% of the searches for non-English

information. The Chi-Square results (Table 24) show that advanced search option was

used differently by different language groups (Pearson Chi-Square (3, N=199) = 8.384, p < .05). So the hypothesis is supported: there are statistically significant differences in users' search behaviors in employing advanced search options by users' background.

		Frequency	Cumulative Percent
Valid	No	52	26.1
	Yes	147	100.0
	Total	199	

Table 24. Chi-Square test of Advanced search option use by users' background

Crosstab of Advanced search option

			native language				
		Chinese Japanese Korean English					
Advanced search option use including limit option	No	9	17	8	18	52	
	Yes	46	34	36	31	147	
Total	=	55	51	44	49	199	

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.384(a)	3	.039
Likelihood Ratio	8.511	3	.037
Linear-by-Linear Association	3.018	1	.082
N of Valid Cases	199		

a 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.50.

1.4.3. Search behavior by researcher vs. library professionals

1.4.3.1 For continuous variables

A single factor or one way ANOVA was conducted to see if there are differences

in subjects' search behaviors by their profession. Among various search behavior

variables, the following three variables show statistically significant differences by two

groups. Library professionals examined more records when they were searching, and

accordingly the search time is longer than researchers. Researchers expressed more difficulties than library professional when they searched for non-English information. So the hypothesis here is supported: there are statistically significant differences in users' IR system use and search behavior by users' background and their experiential knowledge.

The number of examined records: F(1, 197) = 8.905, p < .005

Difficulty index using the system: F(1, 197) = 8.26, p < .005

Search time: *F* (1, 197) = 4.30, *p*<.05

Table 25. ANOVA test for search behaviors by research vs. library professionals

		Sum of Squares	df	Mean Square	F	Sig.
The number of examined records	Between Groups	157.548	1	157.54 8	8.905	.003
	Within Groups	3485.406	197	17.692		
	Total	3642.955	198			
Difficulty index using the system	Between Groups	15.241	1	15.241	8.261	.004
	Within Groups	363.443	197	1.845		
	Total	378.683	198			
Search time (minutes)	Between Groups	78.453	1	78.453	4.298	.039
	Within Groups	3596.230	197	18.255		
	Total	3674.683	198			

Table 26. Descriptive for search behaviors	by research vs. libra	ary professionals
--	-----------------------	-------------------

		Ν	Mean	S. D
The number of examined records	Researcher	98	2.08	1.914
	library professional	101	3.86	5.595
	Total	199	2.98	4.289
Difficulty index using the system	Researcher	98	2.97	1.296
	library professional	101	2.42	1.416
	Total	199	2.69	1.383
Search time (minutes)	Researcher	98	5.05	3.279
	library professional	101	6.31	5.053
	Total	199	5.69	4.308

1.4.3.2. For nominal variables

A Chi-Square test was done for nominal variables to see if there are statistically significant differences for searches using advanced options/Boolean logic for librarians compared to researchers. Boolean logic was used at almost the same level for both researchers and library professionals, so further testing was not needed.

The Chi-Square test show statistically significant differences between library professionals who used advanced search options more often at 82% use rate than researchers at a 65% use rate (Pearson Chi-Square F(1, N=199) = 7.36, p = .007). So the hypothesis is supported: there are statistically significant differences in users' IR system use and search behavior, especially in employing the advanced search options based on users' background.

 Table 27. Chi-Square Tests for advanced option use by researchers vs. library professional

 Crosstab for advanced search option use

		Researcher	Library professional	Total
Advanced search option use	No Yes	34 64	18 83	52 147
Total		98	101	199

Chi-Square test

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.335(b)	1	.007
Continuity Correction(a)	6.487	1	.011
Likelihood Ratio	7.421	1	.006
Linear-by-Linear Association	7.298	1	.007
N of Valid Cases	199		

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 25.61.

1.4.4. Search behavior by experiential knowledge (topic, system, and language knowledge)

Bivariate correlations were run between search behavior variables and experiential knowledge. Non-English searching cases were chosen for this analysis based on the original premises articulated for this hypothesis (N=199). There was not any relationship between search behavior variables and topic/system knowledge in experiential knowledge.

There were statistically significant relationships between the following search behavior variables and the assigned language knowledge for the search:

The number of different languages used: r (199) = .45, p < .001

Search time (minutes): r (199) = .17, p < .05

The number of query changed: r (199) = .30, p < .001

The higher the language knowledge scores the higher the number of different languages searches, the more query changes and the more search time in non-English information searches.

	The number of different language		The number of query changes
Pearson Correlation	.449(**)	.166(*)	.294(**)
Sig. (2-tailed)	.000	.019	.000

Table 28. Correlation of language knowledge with search behaviors

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Multivariate analysis of variance (MANOVA) was used to test for search behavior variables with three experiential knowledge variables (topic, system and language) while assessing comprehensive variance analysis constructed from factorial search behavior variables (first language choice and advanced option use). The following table 29 shows statistically significant findings from this MANOVA analysis.

				Hypot	U		Partial
				hesis			⊂ta Square
Effect		Value	F	df	Error df	Sig.	d
Intercept	Pillai's Trace	.420	44.473(a)	3.000	184.000	.000	.420
	Wilks' Lambda	.580	44.473(a)	3.000	184.000	.000	.420
	Hotelling's Trace	.725	44.473(a)	3.000	184.000	.000	.420
	Roy's Largest Root	.725	44.473(a)	3.000	184.000	.000	.420
numlangchange	Pillai's Trace	.073	4.833(a)	3.000	184.000	.003	.073
	Wilks' Lambda	.927	4.833(a)	3.000	184.000	.003	.073
	Hotelling's Trace	.079	4.833(a)	3.000	184.000	.003	.073
	Roy's Largest Root	.079	4.833(a)	3.000	184.000	.003	.073
examedrecord	Pillai's Trace	.059	3.846(a)	3.000	184.000	.011	.059
	Wilks' Lambda	.941	3.846(a)	3.000	184.000	.011	.059
	Hotelling's Trace	.063	3.846(a)	3.000	184.000	.011	.059
	Roy's Largest Root	.063	3.846(a)	3.000	184.000	.011	.059
difficultyindex	Pillai's Trace	.092	6.204(a)	3.000	184.000	.000	.092
	Wilks' Lambda	.908	6.204(a)	3.000	184.000	.000	.092
	Hotelling's Trace	.101	6.204(a)	3.000	184.000	.000	.092
	Roy's Largest Root	.101	6.204(a)	3.000	184.000	.000	.092

Table 29. Multivariate Tests(c) for search behaviors by experiential knowledge

a Exact statistic

b The statistic is an upper bound on F that yields a lower bound on the significance level.

c Design: Intercept+numterms+numlangchange+examedrecord+difficultyindex+searchtime+advanceuse

These statistically significant findings emerged as the three experiential

knowledge levels differed from the number of language changes, the difficulty index,

number of examined records, and first language choice. This means that there are significant differences among search behaviors where knowledge levels of the task and topic differ. These knowledge levels were coded categorically in the MANOVA analysis and can be regarded as independent factors, even though they may exist on a theoretical continuum. The dependent measures are the three numeric variables: language changes, records examined, and difficulty.

The findings may be significant meaning they did not happen by chance but the effect sizes are not large with approximately six to nine percent of the knowledge level variability/assignment explained by the independent measures. Interactions emerged as significant and explained a large percentage of the knowledge grouping assignment (eta squared = 42%). Such interactions can mean that the model is somewhat confounded due to overlap among the dependent variables; the interactions can also indicate that search behaviors are interdependent when being explained by knowledge level.

Tests of Between-Subjects Effects were then computed to assess the individual differences between variables in the larger data set. As seen in the table 30, topic familiarity does not show overall significant results and it can be considered that it does not account for any effect size explanations (the 6.6% eta squared reported below could have occurred by change since the findings were non-significant). System familiarity shows significant results (F(12, 186) = 2.051, p. = .022) and although its effect size may appear modest (eta squared = 11.7%), this finding did not occur by chance. For a single variable, it achieved a reasonably strong effect size given the complexity of the overall model and the influence of other variables on the search environment including the

eta squared values for the intercept computation).

0	Den en den () (eriekte	Type III Sum of	-14	Mean	F	0.1	Partial Eta
Source	Dependent Variable	Squares	đt	Square	F	Sig.	Squared
Model	I opic familiarity %	13488.781(a)	12	1124.065	1.096	.366	.066
	Assigned language knowledge	136737.409(b)	12	11394.784	10.278	.000	.399
	System familiarity %	19653.016(c)	12	1637.751	2.051	.022	.117
Intercept	Topic familiarity	58499.194	1	58499.194	57.017	.000	.235
	Assigned language knowledge	3603.016	1	3603.016	3.250	.073	.017
	System familiarity	54829.057	1	54829.057	68.657	.000	.270
NumLang	Topic familiarity	76.189	1	76.189	.074	.786	.000
	Assigned language knowledge	15509.619	1	15509.619	13.990	.000	.070
	System familiarity	852.918	1	852.918	1.068	.303	.006
Examrecord	Topic familiarity	2165.956	1	2165.956	2.111	.148	.011
	Assigned language knowledge	8702.366	1	8702.366	7.850	.006	.040
	System familiarity	917.678	1	917.678	1.149	.285	.006
Difficultyindx	Topic familiarity	22.036	1	22.036	.021	.884	.000
	Assigned language knowledge	17531.754	1	17531.754	15.814	.000	.078
	System familiarity %	3100.131	1	3100.131	3.882	.050	.020
Error	Topic familiarity % [0- 100%]	190836.305	186	1026.002			
	Assigned language knowledge	206201.787	186	1108.612			
	System familiarity %	148537.939	186	798.591			
Total	Topic familiarity % [0- 100%]	1056579.000	199				
	Assigned language knowledge	528700.000	199				
	System familiarity %	1229100.000	199				
Corrected Total	Topic familiarity % [0- 100%]	204325.085	198				
	Assigned language knowledge	342939.196	198				
	System familiarity %	168190.955	198				

Table 30. Tests of Between-Subjects Effects for search	n behaviors by experiential knowledge
--	---------------------------------------

a R Squared = .066 (Adjusted R Squared = .006)/ b R Squared = .399 (Adjusted R Squared = .360) / c R Squared = .117 (Adjusted R Squared = .060)

The following variables achieved statistically significant differences with assigned language knowledge when searches are for non-English information.

Number of language changes: F(1, 186) = 14, p < .001 (7 % variability) Examined records: F(1, 186) = 7.9, p < .05 (4 % variability) Difficulty index: F(1, 186) = 15.81, p < .001 (8 % variability)

The total variance explained can be partitioned by dividing the explained Type III Sum of Squares by the Adjusted Total Variance in the model for each major component of the variable breakouts. For example, language knowledge explains 136,737 units of variability in the corrected model where the corrected total number of language units of variability equal 342,939. Thus, this component of the model shows that 39.9% of the language variability explained is accounted for within the model (that is: 136,737 divided by 342,939 = 39.9% of the variance explained). Nonetheless, language itself is not fully explained in terms of the total model but these results still provide a strong argument that language knowledge for a target task has a direct effect on users' search behaviors in a multi-language IR environment. Hence, this has implications for the design of CLIR systems which would be used by those assigned to search in languages not known to them. **R 1.5.** Are their user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine?

H1.5: There are statistically significant differences in users' IR system use and search behaviors for non-English information when compared to English language searching.

A. Differences by English vs. Non-English searches (CJK).

A one way ANOVA test was done with the numeric and continuous search behavior variables.

The number of cases between English searching and non-English searching is not even: English searching (n=89); non-English searching (combined CJK searching n=199). Although the number of searches is not equal, comparisons can still be made if there is homogeneity of variance so that the variability within each set of data is approximately equal or if it can be adjusted for statistical comparisons. Comparison by each language can show if statistical differences exist when comparing English and non-English searching with the following variables: number of different languages, number of query changes, difficulty using system, number of examined records, and search time.

As seen in the descriptive statistical result presented below in the table 31, non-English searching consumes more time and effort than English language searching. A number of plausible explanations could be offered for these patterns. It is interesting that three variables showed distinct descriptive statistical differences between English and non-English searching: the number of different languages, the difficulty index, and the number of examined records. These will now be tested with inferential statistics.

		Ν	Mean	S. D
The number of different language	Eng search	89	.08	.376
	Non-Eng	199	.41	.876
	Total	288	.31	.772
The number of query changed	Eng search	89	2.16	2.641
	Non-Eng	199	2.93	3.114
	Total	288	2.69	2.993
Difficulty index using the system	Eng search	89	.99	1.192
	Non-Eng	199	2.69	1.383
	Total	288	2.16	1.541
The number of examined records	Eng search	89	5.20	9.714
	Non-Eng	199	2.98	4.289
	Total	288	3.67	6.533
Search time	Eng search	89	5.16	3.652
	Non-Eng	199	5.69	4.308
	Total	288	5.52	4.118

Table 31. Descriptive statistics of search behaviors by English vs. non-English searches

The one way ANOVA test result (Table 32) confirms that the three search behaviors are statistically significantly different for non-English vs. English searches: the number of different language used, the number of queries examined and the difficulty index using the system. There were more different languages used, and more records examined in non-English searching than English searching. Non-English searching was more difficult than English searching according to users' responses on the difficulty index for each search. The average number of search terms and search times does not show differences between two searches.

The number of different language used: F(1, 286) = 11.53, p = .001

The number of query examined: F(1, 286) = 7.24, p = .008

Difficulty index using the system: F(1, 286) = 98.30, p < .001

	, [Sum of		Mean		-
		Squares	df	Square	F	Sig.
The average number of search terms	Between Groups	1.352	1	1.352	1.530	.217
	Within Groups	252.634	286	.883		
	Total	253.986	287			
The number of different language	Between Groups	6.632	1	6.632	11.531	.001
	Within Groups	164.480	286	.575		
	Total	171.111	287			
The number of query changed	Between Groups	34.118	1	34.118	3.846	.051
C C	Within Groups	2536.993	286	8.871		
	Total	2571.111	287			
The number of examined records	Between Groups	302.349	1	302.349	7.238	.008
	Within Groups	11947.314	286	41.774		
	Total	12249.663	287			
Difficulty index using the system	Between Groups	174.275	1	174.275	98.298	.001
	Within Groups	507.055	286	1.773		
	Total	681.330	287			
Search time (minutes)	Between Groups	23.067	1	23.067	1.362	.244
	Within Groups	4842.763	286	16.933		
	Total	4865.830	287			

Table 32. ANOVA test for search behaviors by English vs. non-English searches

B. differences by each language search (CJKE)

In a one way ANOVA test, the number of different languages used and the 'difficulty index' are statistically different for the four different language searches. The number of different languages used resulted in significant results: F(3, 284) = 5.25, p <.005. Post hoc comparison tests shows significant differences between English vs. Chinese .48(*) for their mean differences at p < .05 probability level. In English searching, there were fewer languages used than Chinese and Japanese searching.

The overall difficulty index level is significant at F(3, 284) = 34.41, p < .0001, and its post hoc test shows: English vs. Chinese -1.77 (*); English vs. Japanese -1.50 (*); and, English vs. Korean -1.82 (*). Each CJK language search shows it was more difficult than English searches. So the hypothesis that there are statistically significant differences in users' search patterns for non-English information when compared to English language searching is supported. The effect size for this ANOVA can be stated by comparing the "explained" between sum of squares" with the total variability in the model: (eta squared can be expressed as 8.99 + 181.6 divided by 681.3 = 28% of the total variance is explained by language differences.

Table 33. ANOVA test of search behaviors by different language searches

		Sum of Squares	df	Mean Square	F	Sig.
The number of different language	Between Groups	8.993	3	2.998	5.251	.002
	Within Groups	162.118	284	.571		
Difficulty index using the system	Between Groups	181.616	3	60.539	34.406	.000
0 7	Within Groups	499.714	284	1.760		
	Total	681.330	287			

		(J)	Mean	
Dependent Variable	(I) Language	Language	Difference (I-J)	Sig.
The number of different language used	Chinese	Japanese	.17	
		Korean	.27	1.000
		English	.48(*)	.267
	Japanese	Chinese	17	.001
		Korean	.10	1.000
		English	.31	1.000
	Korean	Chinese	27	.071
		Japanese	10	.267
		English	.21	1.000
	English	Chinese	48(*)	.495
		Japanese	31	.001
		Korean	21	.071
	-	-	-	

Multiple Comparisons of search behaviors by different language searches (Bonferroni)

Difficulty index using the system	Chinese	Japanese	.27	.495
		Korean	05	1.000
		English	1.77(*)	1.000
	Japanese	Chinese	27	.000
		Korean	32	1.000
		English	1.50(*)	.971
	Korean	Chinese	.05	.000
		Japanese	.32	1.000
		English	1.82(*)	.971
	English	Chinese	-1.77(*)	.000
		Japanese	-1.50(*)	.000
		Korean	-1.82(*)	.000
			.17	.000

* The mean difference is significant at the .05 level.

		Ν	Mean	Std. Deviation
The number of different language	Chinese	63	.56	1.028
	Japanese	67	.39	.816
	Korean	69	.29	.769
	English	89	.08	.376
	Total	288	.31	.772
Difficulty index using the system	Chinese	63	2.76	1.266
	Japanese	67	2.49	1.418
	Korean	69	2.81	1.448
	English	89	.99	1.192
	Total	288	2.16	1.541

Descriptives of search behaviors by different language searches

C. System evaluation by English and non-English search

A one way ANOVA test was done for system evaluation for English and non-English searches. The results reveal that there are statistically significantly differences in users' evaluations of English searching and non-English searching with F(1, 286) =95.64, p < .001. English searching shows a much higher satisfaction score for searching than non-English searching with over a 100% difference between their respective means. This model explains 25% of the dependent measure's variability.

Table 34. ANOVA test of Satisfaction of the search result Descriptives of Satisfaction of the search result % [0-100%]

		L	
	N	Mean	S. D
English searching	89	82.13	18.767
Non-English searching	199	40.41	38.213
Total	288	53.31	38.580

ANOVA test of Satisfaction of the search result % [0-100%]

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	107052.518	1	107052.518	95.644	.001
Within Groups	320114.593	286	1119.282		
Total	427167.111	287			

The following box and whisker plot is showing satisfaction distributions for

English vs. non-English searching.



Figure 4. Box and whisker plot of satisfaction by English vs. non-English searches

5.2. Research question 2

- What issues are present when individuals search non-English information via online DBs, OPACs and web search engine?
 - 2.1 What kinds of issues and limitations exist when searching non-English information?
 - 2.2 How do those factors affect the access to information contained in such systems?H2.2.1: Users will avoid databases if the system lacks efficient andcomprehensive language coverage when they look for non-English information.

H2.2.2: If the system lacks efficient and comprehensive language coverage, then users' search time will be shorter than with other systems when they search for non-English information.

H2.2.3: Users have better knowledge about scholarly academic journals in their areas written in English than for journals written in non-English.

R2. What issues are present when individuals search non-English information via online DBs, OPACs and web search engine?

2.1. What kinds of issues and limitations exist when searching for non-English information?

A. Issues in accessing non-English information

Survey and experiment participants provided the following issues when they access and use non-English information. These issues are very similar to the users' expectations presented in the first research question. Some representative comments were quoted.

14010 001 155405 11	
Category	Comments
Lack of non-	Participants expressed there are difficulties and limitations to type
English access	and search in their target languages. This issue reinforces the need for
via indexing	a cross language information retrieval function in such systems.
terms	"Most systems do not allow to type in Korean (target language)";
	"It is still impossible to search in Japanese or Chinese on English-
	language library catalogs, and too often the characters in the entries
	(when they are there at all) have display problems or cannot be
	included in emailed/downloaded records."
Lack of non-	Many participants complained about lack of non-English material
English records	coverage in current online database system.
in major online	"No online databases for Chinese articles in the university
databases which	electronic databases":
index journals	"Many databases do not yet include Asian character information
J	(as, for instance WorldCat does), making searches and interpreting
	results imprecise and time-consuming. Most US academic databases
	(esp. journals) do not include non-English datasets, therefore
	searches must be repeated across many separate databases
	worldwide".
	"I need to access primary and secondary sources for my research
	on pre-modern Chinese history more or less constantly while engaged
	in research. I have limited (actually no) access to such sources
	through my home institution which is a small college "
	anough my nome institution, which is a small conege.
Lack of English	Users often encountered a bibliographic record without abstract or
translation of	summary. It is hard to judge whether that record would be relevant
abstract and	for them with only reading the title and, if available, index term
summarv	descriptors. If the bibliographic record is written using a Romanized
j	script (i.e., a transliterated record), it can be no more than mere
	symbols rather than content conveying information for those who do
	not know the language or how the language is pronounced using a
	narticular dialect. The following comments were made from
	experiment participants
	"if there was no abstract or a very limited abstract it was difficult
	or impossible to determine whether the article would be useful"
	(Subject #3)
	"I can't easily interpret the relevance of the retrieved results since
	the titles of the retrieved books did not say clearly about the content"
	(Subject #28 - task done with WorldCat))
	"I don't understand the Jananese title And there was no English
	abstract or description about the book" (Subject #20)
	"I ack of English information (in the record) made you wonder if
	Lack of Elightin mormation (in the fection) made you wonder if

Table 35. Issues in accessing non-English information

	you are doing the right thing" (Subject # 15)
Lack of	Romanization was introduced in order to transcribe non-Roman
coherent and	alphabet letters to Roman letters. Since many non-Roman letters do
understandable	not have exactly equivalent Roman letters for transliteration to occur,
access to non-	this Romanization system is destined to have inherent problems by its
Roman	very nature. Romanization issues are summarized in a separate table
language	below.
materials	

Romanization issues from survey participants were summarized as follows:

Category	Comments
Standardization	-"Problems with inconsistent Romanization: Wade-Giles vs. pinyin
in	for Chinese (some libraries have converted records incompletely);
Romanization	Japanese phrase boundaries and use of macrons for long vowels."
in each	- "What's the standard Romanized spelling?"
language	- "One has to enter several queries with different spellings or word
	spacing to do a search that would only require one query if it were in $\sum_{i=1}^{n} \frac{1}{i} + \frac{1}{i}$
	English.
	- The OSCAR database at Ohio State uses a mixture of
	romanization styles for Japanese which is time-consuming when trying to complete a search."
	- "For Japanese, the differences in romanization can be an issue.
	Sometimes Modified Hepburn is used, sometimes a different
	orthographic style. If you can use both, then you have a better
	chance of getting more correct hits."
Difficult to	"There are many ways to Romanize words from languages not
type/to know	written in the English alphabet. If you spell your search entry
Romanized	differently from how it is spelled in the database, it may not always
word	retrieve what you were looking for. This is especially for when you
	do not know a language well and may not know all the possible
	ways to Romanize your search query."
	- "If I do a query using Romanization of Japanese, I often run into
	trouble with long vowels or alternate spelling. Therefore, I have to
	think of what would be the best way to spell a Romanized Japanese
	word in order to get the best results. Also, if I use Japanese to search
	for a person's name, I have to determine what form I should use
	(kanji, hiragana, using the "no" between first and last names, etc)."

Table 36. Romanization issues

	 "I don't (or rarely) understand the title itself. Later I noticed there was a translated title (in English) but it's not enough to understand what the book is about." "I had difficulties in finding Korean novels at Rutgers by means of the Rutgers library catalog because the English-translated titles were rarely readable." "If I have to search in romanized forms, it can be difficult to know how to spell the words as there is not always a standardized way to do so."
Inaccurate data	 "WorldCat: search interface tricky to use and results often imprecise" "WorldCat: many records are duplicated or of poor quality, and many items held by U.S. libraries are not represented, especially non-English materialsitems held outside North America and Western Europe are very poorly represented" "WorldCat: Multiple entries for same item can be difficult to search because it seems very few people who manage the materials in the libraries and are responsible for cataloguing the non-English material understand the language of the material, some materials have very poor descriptions in the databases. For example, material such as anthologies of fiction writers may only state what volume number in the set that it is, and if you are looking for a particular author's works inside that set of material, the library database will tell you that the library does not have what you're looking for, even though it IS actually there." "Some library catalogs don't display the non-English characters correctly." "the records for CJK titles before the late 20th century often have no dates, author names or publisher's information"

B. From observation of the experiment participants' tasks

	1 5 5
Systems	Observation notes
WorldCat	There were apparent differences to use the system between
	experienced and inexperienced users. Most librarians who are
	native CJK speakers were used to the system and knew exactly
	how to use it, whereas, first time users including librarians
	expressed difficulties in using the system.
	Most users were not sure how to start typing their queries

Table 37. Observation note of experiment subjects' system use

	looking for non-English information. Most users typed in English and most of times they did not get good result. For example, subject #27 said "I am so frustrated it seems it has some but I don't get any results. Is this because I did the search in English? How do I know equivalent word for that?" Even CJK native subjects seem to have difficulties in finding information in their own language from the query formulation and reading the bibliographic records. Many CJK subjects tried their own native language and Romanized words for their queries but the number of retrieved information hits were very small and still many records had only Romanized descriptions without their own language description or English translation. Most subjects, especially those who did not know about WorldCat before, were surprised the system has such a huge collection of non-English materials and that it allowed one to search in their own languages. But at the same time, they claimed they want to have a batter index system which allows accessing
	more information in the collection with any language they want.
EBSCO Host	 More information in the collection with any language they want. Although it has many good qualities with databases in various subjects, it does not appear easy to use. First of all, depending on the database, it varies whether the specific database has non-English materials or not. Most databases do not seem to have many non-English collections. Only one expert subject effectively used an advance search option called 'limit' which allows users to find specific language collections. Also, even though they found non-English information, many records do not provide English translation of the description from title to abstract. Romanized descriptions prevailed and subjects could not decide whether retrieved information was relevant or not. Also subjects were surprised it does not allow one to type their query input in languages other than English. The language box on the first screen seems to allow users to type and read the languages, but it only allows them to change the interface. Most subjects also expressed inconvenience with the limited query input box which only allows one word in each box.
Google	Most subjects have much experience with Google searching and almost everyone talked about the quality of the information. Even if they use Google scholar, a number of subjects indicated that the retrieved information might not be from an authentic information source. Most subjects including native CJK subjects were surprised that Google has a language tool where they can translate the query and retrieve information in language they want to Also
they seemed happy to have the translation website option. For	
--	
English speakers including librarians, they appeared not to care	
about the language service but they were still amazed with what	
they can do with it. But most subjects noted the poor quality of	
the translations provided from Google.	

2.2. How do those factors affect the access to information contained in such systems?

H 2.2.1: Users will avoid databases if the system lacks efficient and

comprehensive language coverage when they access non-English information.

A. verifying an inefficient system by comparison of the three systems

As seen the following descriptive table, the Google system has the highest efficiency score (M=66.67, SD=27.26), followed by WorldCat (M=46.03, SD=40.62), and then trailed by EBSCO Host (M=8.26, SD=20.46). The satisfaction and difficulty index scores also indicate that the EBSCO Host system is the most difficult to use and this is found in its low satisfaction score. The WorldCat system (M=3.60, SD=2.52) showed the highest number of query changes, whereas Google (M=1.90, SD=2.86) shows the fewest changes. So it can be said the EBSCO Host was regarded by users as the most inefficient system among three systems when searching for non-English information.

		Ν	Mean	S. D.	Std. Error
The number of query changes	WorldCat	65	3.60	2.524	.313
	EbscoHost	67	3.03	3.393	.415
	Google	60	1.90	2.862	.370
	Total	192	2.87	3.022	.218
The number of examined records	WorldCat	65	3.55	6.083	.754
	EbscoHost	67	1.76	3.051	.373

	Google	60	3.53	2.671	.345
	Total	192	2.92	4.306	.311
Difficulty index using the system	WorldCat	65	2.35	1.178	.146
	EbscoHost	67	3.78	1.071	.131
	Google	60	1.92	1.139	.147
	Total	192	2.71	1.379	.100
Search time (minutes)	WorldCat	65	6.31	4.290	.532
	EbscoHost	67	4.57	3.811	.466
	Google	60	6.18	4.188	.541
	Total	192	5.66	4.153	.300
Record understand level	WorldCat	55	70.45	29.080	3.921
	EbscoHost	26	72.50	30.438	5.969
	Google	55	63.69	30.138	4.064
	Total	136	68.11	29.784	2.554
Satisfaction of the search result	WorldCat	65	48.15	37.910	4.702
	EbscoHost	67	13.73	29.444	3.597
	Google	60	59.28	31.101	4.015
	Total	192	39.62	38.226	2.759
Relevance of intention to use	WorldCat	32	74.06	25.886	4.576
	EbscoHost	10	64.50	32.011	10.123
	Google	33	68.48	26.085	4.541
	Total	75	70.33	26.678	3.081
Efficiency of the system	WorldCat	63	46.03	40.612	5.117
	EbscoHost	65	8.26	20.460	2.538
	Google	58	66.66	27.255	3.579
	Total	186	39.26	38.922	2.854

An ANOVA test was conducted to see whether there are different search and system use behaviors and evaluation by the three different systems (WorldCat, EBSCOHost, and Google). These results encompass which system does not support users' non-English searching. It was confirmed that there are statistically significant differences for the following variables by the three systems using the ANOVA Post Hoc test to reveal specific details (see Table 39). Note that the EBSCOhost system received lower evaluation scores than other two systems. So it is reconfirmed that the EBSCO Host is the least regarded system among the three systems searching for non-English information. compared across the three systems.

The number of query changes: F(2, 189) = 5.312, p < .05

The number of examined records: F(2, 189) = 3.85, p < .05

Difficulty index using the system: F(2, 278) = 25.52, p < .0001

Search time (minutes): *F* (2, 278) = 3.69, *p* < .05

Satisfaction of the search result: F(2, 278) = 33.37, p < .0001

Efficiency of the system: *F* (2,289) =58.13, *p*<.0001

		Sum of Squares	df	Mean Square	F	Sia
The number of query	Between Groups	92.804	2	46.402	5.312	.006
changed	Within Groups	1650.940	189	8.735		
	Total	1743.745	191			
The number of examined records	Between Groups	138.654	2	69.327	3.850	.023
	Within Groups	3403.174	189	18.006		
	Total	3541.828	191			
Difficulty index using the system	Between Groups	122.158	2	61.079	47.883	.000
	Within Groups	241.087	189	1.276		
	Total	363.245	191			
Search time (minutes)	Between Groups	123.718	2	61.859	3.687	.027
	Within Groups	3171.277	189	16.779		
	Total	3294.995	191			
Satisfaction of the search result % [0-100%]	Between Groups	72837.436	2	36418.71 8	33.371	.000
	Within Groups	206263.809	189	1091.343		
	Total	279101.245	191			
Efficiency of the system	Between Groups	108876.498	2	54438.24 9	58.126	.000
	Within Groups	171389.594	183	936.555		
	Total	280266.091	185			

Table 39. ANOVA tests for search behaviors by three systems

			Mean		
			Difference (I-	Std.	
Dependent Variable	(I) System	(J) System	J)	Error	Sig.
The number of query changed	WorldCat	EbscoHost	.570	.515	.808
		Google	1.700(*)	.529	.005
	EbscoHost	WorldCat	570	.515	.808.
		Google	1.130	.525	.098
	Google	WorldCat	-1.700(*)	.529	.005
		EbscoHost	-1.130	.525	.098
The number of examined records	WorldCat	EbscoHost	1.793(*)	.739	.049
		Google	.021	.760	1.000
	EbscoHost	WorldCat	-1.793(*)	.739	.049
		Google	-1.772	.754	.059
	Google	WorldCat	021	.760	1.000
		EbscoHost	1.772	.754	.059
Difficulty index using the system	WorldCat	EbscoHost	-1.422(*)	.197	.000
		Google	.437	.202	.096
	EbscoHost	WorldCat	1.422(*)	.197	.000
	. .	Google	1.859(*)	.201	.000
	Google	WorldCat	437	.202	.096
		EbscoHost	-1.859(*)	.201	.000
Search time (minutes)	WorldCat	EbscoHost	1.741(*)	.713	.047
		Google	.124	.733	1.000
	EbscoHost	WorldCat	-1.741(*)	.713	.047
		Google	-1.616	.728	.083
	Google	WorldCat	124	.733	1.000
		EbscoHost	1.616	.728	.083
Satisfaction of the search result	WorldCat	EbscoHost	34.423(*)	5.751	.000
		Google	-11.129	5.914	.184
	EbscoHost	WorldCat	-34.423(*)	5.751	.000
	Caarla	Google	-45.552(*)	5.872	.000
	Google		11.129	5.914	.184
		EDSCOHOST	45.552(*)	5.872	.000
Efficiency of the system	worldCat	EDSCOHOSt	37.770(*)	5.411	.000
	Ebaaallaat	Google	-20.623(*)	5.569	.001
		Coords	-3/.//U(*)	5.411	.000
	Casala	Google	-58.394(*)	5.528	.000
	Google	vvoridCat	2U.623(^)	5.569	.001
			30.394(*)	ວ.ວ∠୪	.000

Post Hoc tests for search behaviors by three systems (Bonferroni)

* The mean difference is significant at the .05 level.



three systems

Figure 6. Satisfaction of the search by three systems

B. System choice

Participants in the experiment (N=32) were asked to choose any IR system to search for non-English information. EBSCOhost was avoided more than other DBs. An overwhelming majority, 69% (n= 22 people) chose Google to search for non-English information, whereas WorldCat and Yahoo were chosen by only 9.4% of the subjects (3 people). Individual participation (n=1 person each) chose Medline, Web of Science, Naver (Korean web search engine) and Wikipedia. This result shows that when individuals need to search for non-English information, not many of them use or know about either online and journal databases or WorldCat. The hypothesis is supported: users will avoid databases if the system lacks efficient and comprehensive language coverage when they access non-English information.

H2.2.2: If the system lacks efficient and comprehensive language coverage, then users' search time will be shorter than with other systems when they search for non-English information.

A. A one way ANOVA was used to assess the differences in search time by the three systems. From previous test results, EBSCO Host had the lowest system efficiency score when searching for non-English information. At the descriptive table shows, the EBSCOHost system was involved in the shortest search times for non-English information.

Table 40. Descriptives of Search time by three systems Ν Mean S. D WorldCat 4.290 65 6.31 EbscoHost 67 4.57 3.811

60

192

Google

Total

The ANOVA test confirmed that three systems has statistically different search times (F(2, 189) = 3.69, p < .05). The post hoc tests compare all three means with each other in pairs. This resulted in a finding that there is a statistically significant difference between EBSCO Host and WorldCat. There was no significant difference in search times between the other pairs of scores. In fact, it was surprising that there was not a statistically significant difference between EBSCO Host times and Google times. Further, the effect size for this overall F test was weak: eta squared only accounts for a 3.8%

6.18

5.66

4.188

4.153

explanation level for the variability in search times across systems. Thus, the hypothesis is partly and weakly supported: If the system lacks efficient and comprehensive language coverage, then users' search time will be shorter than with other systems when they search for non-English information. But this result is fragile given the anemic effect size.

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	123.718	2	61.859	3.687	.027
Within Groups	3171.277	189	16.779		
Total	3294.995	191			

Table 41. ANOVA test of Search time (minutes) by three systems

		Mean			95% Confide	ence Interval
		Difference				
System	(J) System	(I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
WorldCat	EbscoHost	1.741(*)	.713	.047	.02	3.46
	Google	.124	.733	1.000	-1.65	1.90
EbscoHost	WorldCat	-1.741(*)	.713	.047	-3.46	02
	Google	-1.616	.728	.083	-3.37	.14
Google	WorldCat	124	.733	1.000	-1.90	1.65
	EbscoHost	1.616	.728	.083	14	3.37

Post hoc tests of Search time (minutes) by three systems (Bonferroni)

* The mean difference is significant at the .05 level.

B. Cross-tab and Chi-Square test for when the user 'gives-up' a search in the three systems

This hypothesis was confirmed concerning when an individual voluntarily ends or "gives up" a search. A Chi-Square test was done for those giving up or abandoning their search for each of the three systems. As seen below (N=199), many give-up their non-English searching when using the EBSCOHost system (Give-up: n=50) compared to other systems (WorldCat: n=16; Google: n=4). So this hypothesis is supported.

The Pearson Chi-Square test result shows there is statistically difference for the

three systems for users who give-up their search X^2 (3, N = 199) = 71.16, p < .001.

crosstabulation of give-up by three systems							
System							
	WorldCat EbscoHost Google others					Total	
Give	No	49	17	56	6	128	
up	Yes	16	50	4	1	71	
Total		65	67	60	7	199	

Table 42. Chi-Square Tests of give-up by three systems Crosstabulation of give-up by three systems

Chi-Square Tests of give-up by three systems

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	71.157(a)	3	.000
Likelihood Ratio	75.735	3	.000
Linear-by-Linear Association	4.959	1	.026
N of Valid Cases	199		

a 2 cells (25.0%) have expected count less than 5. The minimum expected count is 2.50.

The following figure depicts this difference.



Bar Chart

Figure 7. Termination of search for the three systems

H 2.2.3: Users have better knowledge about scholarly academic journals in

their areas written in English than journals written in non-English.

Out of 204 participants, 112 individuals provided well known journal names in

their study or work field. Among the 112 respondents only 44 people recalled or

mentioned the non-English journal names in their area. The survey participants' were

asked the reason why major journals in their area are primarily in English. Selective

representative comments are summarized below:

- 1. English is a common language
 - "most readers can read English"
 - "English is the predominant international world language"
- 2. English is the major language in some research area
 - "certain research area has the tendency of being in written English"
 - "the mainstream research in this area has the tendency of being written in English."
 - "anthropology remains a western-dominated discipline"

3. Some research areas have dominance in the US

- "no advanced study in Japan LIS"
- "American universities dominate in the study of political science globally "
- 4. Need to respect other language works
 - "US universities in this field don't give high respect for other languages"
- 5. Accessibility
 - "many works are not accessible from major English database"
 - "there might be some journals but don't know what are they because I've studied in the states for master's and my doctorate degree."

6.3. Research question 3:

3. How do bibliographic records facilitate or hinder the understanding of bibliographic and retrieved information?

H3.1: There will be statistically significant differences for users' understanding of full records which includes English translation, Romanized and vernacular languages than for other record displays.

H3.2: There will be statistically significant differences in bibliographic

understanding level between English searching and non-English searching.

H3.3: There will be statistically significant differences for bibliographic

understanding between researchers and library professionals.

R 3. How do bibliographic records facilitate or hinder the understanding of

bibliographic and retrieved information?

A linear regression model revealed that four variables explain 46% of the variability in the degree of understanding the bibliographic record, R (288) = .676, F (4, 227) = 47.66, p < .0001.

Table 43. Regression test for bibliographic record understand level

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.676(a)	.456	.447	20.109

a Predictors: (Constant), Topic familiarity % [0-100%], Assigned language knowledge, job experience, Satisfaction of the search result % [0-100%]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regressio n	77089.334	4	19272.333	47.658	.001(a)
	Residual	91796.649	227	404.391		
	Total	168885.983	231			

ANOVA(b) value for regression of bibliographic record understand level

a Predictors: (Constant), Topic familiarity % [0-100%], Assigned language knowledge, job experience, Satisfaction of the search result % [0-100%]

b Dependent Variable: Record understand level % [0-100%]

The regression model equation can be expressed as follows:

Record understanding level = .462 Satisfaction of the search result

+ .299 Assigned language knowledge

+ .215 job experience

+ .191 Topic familiarity

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		В	Std. Error	Beta		
1	(Constant)	25.498	4.213		6.052	.000
	Language knowledge	.179	.031	.299	5.830	.000
	Satisfaction of the search result %	.382	.043	.462	8.996	.000
	job experience	.800	.183	.215	4.368	.000
	Topic familiarity %	.157	.041	.191	3.854	.000

Table 44. Coefficients(a) of each variable for explaining bibliographic record understand level

a Dependent Variable: Record understand level % [0-100%]

This result indicates that when users are satisfied that their query obtained satisfactory results and if they know the appropriate language, have experience and topic familiarity, then the person will give high understanding scores to the bibliographic record. In other words, the search environment probably needs to be clear to the user from start to finish. Confusion can result from several key variables, such as during query construction or from lack of familiarity with the topic. It is assumed that when confusion enters the search process stream of events, then the user will have more difficulty understanding the IR results retrieved.

H3.1: There will be statistically significant differences in users' understanding levels with full records which include English translation, Romanized and vernacular language compared to other record displays.

A. In the online survey, participants (N=171) were asked to evaluate their understanding levels of four different bibliographic records from the WorldCat system. The four records provide a sample entry for Chinese, Japanese, and Korean materials. The results show that people can understand the bibliographic records when the record has either been described in English only (below, 4 in the table 45) or Romanized, original language (vernacular) and English translation (3 in the table 45). It was hypothesized that users need a full descriptive record which includes English abstract and translations of key entities such as title, descriptors and vernacular language. This hypothesis was not supported.

	1. Romanized	2. Romanized	3. Romanized	4. English only record				
	bibliographic	and vernacular	Vernacular and	(with English				
	record	record	English translation	descriptor)				
Mean	3.94	4.64	5.98	5.95				

Table 45. Mean of four bibliographic records (N=171)

B. Users' needs and comments about 4 different bibliographic records

Survey respondents were asked what features should be included in each

bibliographic record.

1. Only romanized record

- Need English translations accompanying each details
- CJK character display
- Keywords; the descriptor field is insufficient
- Meaning of book title/ brief description of title's contents (plot summaries, themes)/ story titles more specific contents info for example like chapter by chapter
- Description of textual format.
- 2. Vernacular language (original character) + romanized bibliographic record
 - Translation of title, include description in English
 - Keywords / subjects field is missing in Chinese as well as in English
 - Description needed (e.g., abstract) / Summary of the plot
 - Link to records of other editions of this book available / "Since many of Zhang Ailing's novels have been published in English under the name Eileen Chang it would be good to have cross references to her usual English rendering and it would be good to have a cross reference to the translation /
 - "If necessary, links to other Romanizations of same author and title (Wade Giles) or simplified Chinese characters."
 - "adding tones to the Chinese language pinyin would be most helpful for how to pronounce some of the words"
- 3. English translation with vernacular and Romanized versions
 - "Link everything (English/Romanization alongside non-Roman letters), tells you what the book is about and what language it's in. I have a good idea of whether this could be useful to me just by looking at this record "
 - TOC (table of content)
 - Descriptor field is insufficient
 - A brief description (e.g. abstract) about the contents in Japanese language
 - English record to accompany Chinese one
 - Need to see the vernacular: "Chinese language has so many homophones that romanization is confusing to me"
 - Link to record for original publication (with Chinese characters for title)

H3.2: There will be statistically significant differences in bibliographic understanding level for English searching compared to non-English searching.

A. One way ANOVA test of bibliographic understanding level for English vs.

non-English searches.

The results reveal statistically significantly differences in subjects' bibliographic understanding level for English searching compared to non-English searching as F(1, 230) = 34.46, p < .001. English results show a much higher understanding of the record than non-English searching — a 30% increase in understanding.

Table 46. ANOVA test for Record understanding level by English vs. non-English search Descriptives of Record understand level

	Ν	Mean	S. D
English searching	89	88.34	15.234
Non-English searching	143	68.31	29.842
Total	232	75.99	27.039

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	22007.634	1	22007.634	34.462	.001
Within Groups	146878.349	230	638.602		
Total	168885.983	231			

ANOVA test of Record understanding level

B. One way ANOVA test was conducted for each language searched

The results show that the record understanding level is statistically significantly

different for different language searches F(3, 228) = 11.87, p < .0001.

As seen below a Bonferroni post hoc test by was done and it revealed that each

CJK language searching is different than English searching at the .05 level.

Bibliographic records understand level: English/Chinese: 20.378(*)

English/Japanese: 17.222(*)

English/Korean: 23.099(*)

Table 47.	ANOVA t	ests for	Record	understand	level b	v language	search
						,	

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	22819.250	3	7606.417	11.873	.001
Within Groups	146066.733	228	640.644		
Total	168885.983	231			

Post Hoc test for Record understand level % [0-100%] (Bonferroni)

		Mean				
(I) Language	(J) Language	Difference (I-J)	Std. Error	Sig.	95% Confide	ence Interval
					Lower	Upper
					Bound	Bound
Chinese	Japanese	-3.156	5.039	1.000	-16.57	10.26
	Korean	2.721	5.322	1.000	-11.44	16.89
	English	-20.378(*)	4.503	.000	-32.36	-8.39
Japanese	Chinese	3.156	5.039	1.000	-10.26	16.57
	Korean	5.877	5.251	1.000	-8.10	19.85
	English	-17.222(*)	4.418	.001	-28.98	-5.46
Korean	Chinese	-2.721	5.322	1.000	-16.89	11.44
	Japanese	-5.877	5.251	1.000	-19.85	8.10
	English	-23.099(*)	4.738	.000	-35.71	-10.49
English	Chinese	20.378(*)	4.503	.000	8.39	32.36
	Japanese	17.222(*)	4.418	.001	5.46	28.98
	Korean	23.099(*)	4.738	.000	10.49	35.71

* The mean difference is significant at the .05 level.

The following figures show the differences in bibliographic understanding level for non-English vs. English searching and by each language.



Figure 8. Record understand level by different language search

Figure 9. Record understand level by English vs. non-English search

H3.3: There will be statistically significant differences in bibliographic understanding level by researchers compared to library professionals.

This hypothesis is supported. This hypothesis was made based on an assumption that experienced library professionals will understand non-English bibliographic records better than researchers. With survey data, multivariate analysis of variance (MANOVA, or multivariate GLM) was used to see if means for a set of four types of bibliographic record understanding levels vary by researchers vs. library professionals. Overall test results were statistically significantly different by the two user groups with 11% of the variability explained (Eta Squared = .107). Wilks' Lambda (.893) and other test results yield significant findings at F(4, 283) = .103, p < .001. So the hypothesis is supported although it is not surprising that professional librarians would be able to parse and understand bibliographic records better than researchers.

As seen in the below table (tests of between-subject effects), two variables show significant differences between the two user groups: 'Romanized Vernacular and English translation record understanding level' and 'English only record (with English descriptor) understanding level.' These findings are consistent with the sample who were expected to know English. But it also confirms the value of enhancing records and making English available even to those who know other languages.

Effect		Value	F	Hyp df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.961	1734.786(a)	4.000	283.000	.000	.961
	Wilks' Lambda	.039	1734.786(a)	4.000	283.000	.000	.961
	Hotelling's Trace	24.520	1734.786(a)	4.000	283.000	.000	.961
	Roy's Largest Root	24.520	1734.786(a)	4.000	283.000	.000	.961
reslib	Pillai's Trace	.107	8.501(a)	4.000	283.000	.000	.107
	Wilks' Lambda	.893	8.501(a)	4.000	283.000	.000	.107
	Hotelling's Trace	.120	8.501(a)	4.000	283.000	.000	.107
	Roy's Largest Root	.120	8.501(a)	4.000	283.000	.000	.107

Table 48. Multivariate Tests of for bibliographic understanding level by researchers vs. library professionals

a Exact statistic b Design: Intercept+reslib

Lests of Between-Subjects Effe	cts
--------------------------------	-----

Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Romanized, Vernacular and English translation record understand level	55.125	1	55.125	22.069	.001	.072
English only record (with English descriptor) understand level	28.125	1	28.125	11.120	.001	.037

The descriptive statistics seen below (table 49) reaffirms that library professionals understood the bibliographic records better than researchers. Researchers have a better understating with English translation only (without original language transcript) than with the full record.

Researcher vs Library professional Ν Mean SD Romanized bibliographic Researcher 2.81 1.429 144 record understand level library professional 3.06 1.253 144 Total 2.94 1.347 288 Romanized and vernacular Researcher 4.13 2.183 144 record understand level library professional 4.31 1.799 144 Total 1.999 288 4.22 Romanized, vernacular and Researcher 4.69 1.575 144 English translation record library professional 5.56 1.585 144 Total 5.13 1.637 288 English only record Researcher 5.31 1.575 144 library professional 5.94 1.605 144 Total 5.62 1.618 288

Table 49. Descriptive Statistics for bibliographic understanding level by researchers vs. library professionals

The following figure 10 depicts about the distributions of four bibliographic understanding levels for researcher vs. library professional. It shows well the understanding level difference by researchers and library professionals.



Figure 10. Distributions of four bibliographic understanding levels for researcher vs. library professional

5.4. Research question 4:

4. What explains non-English information users' relevance certainty judgments (intention to use retrieved information)? How do these intentions compare to those reported for English or monolingual users?

H4: There is a statistically significant difference in relevance certainty judgments for searches seeking non-English information compared to searches seeking English information.

R 4. What explains non-English information users' relevance certainty judgments (intention to use retrieved information)? How do these intentions compare to those reported for English or monolingual users?

A logistic regression model revealed that three variables explained the equivalent of 90% (Nagelkerke R Square = .896) of the variability in decisions whether to use the retrieved record or not when searching for non-English information (N=199):

- Relevance
- Number of examined record
- Difficulty index

Table 50. A logistic regression test for intention to use of bibliographic record Omnibus Tests of Model Coefficients for a logistic regression test of intention use

		Chi-square	df	Sig.
Step	Step	4.165	1	.041
	Block	209.142	3	.000
	Model	209.142	3	.000

Model Summary of a logistic regression test of intention to use

		Cox & Snell R	Nagelkerke R
Step	-2 Log likelihood	Square	Square
3	49.636(a)	.662	.896

a Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Variables in the Equation of a logistic regression test of intention to use

		В	S.E.	Wald	df	Sig.	Exp(B)
Step	Examined record	.251	.096	6.862	1	.009	1.285
	Difficulty index	611	.308	3.928	1	.047	.543
	Relevance	.115	.022	26.191	1	.000	1.122
	Constant	- 2.262	.899	6.331	1	.012	.104

a Variable(s) entered on step 1: relevance.

b Variable(s) entered on step 2: examedrecord.

c Variable(s) entered on step 3: difficultyindex.

				Predicted	
Observed		Intention to use the retrieved Print information		Percentage Correct	
			No	Yes	
Step 2	Intention to use the retrieved information	No	116	1	99.1
		Yes	5	71	93.4
	Overall Percentage				96.9

Classification Table(a) of a logistic regression test of intention to use

a The cut value is .500

This model correctly predicted 97% of the decisions to intend to use the record (correctly predicted for **no** use: 99 %; and, for **yes**: 93%). This decision model was based on three salient variables which predicted intended use: relevance judgment, examination

of the record, and difficulty index assessment. Thus, this model may appear obvious and only of use for very short term predictions since those who examine and understand the bibliographic description are more likely to then seek the information it references. There are other variables that affect intention to use a specific document reference by a bibliographic record and these include language knowledge and system knowledge. The contribution of these variables was diminished by the three strong variables in the model. The problem here could be one of multicollinarity (or variable redundancy) or a flawed theoretical model. If it is the latter, then the decision to use the document would need to be recast as an independent variable or it would need to be made continuous and not dichotomous.

H4: There is a statistically significant difference in relevance certainty judgments for searches seeking non-English information compared to searches seeking English information.

One way ANOVA by English vs. non-English searching is used to test this hypothesis. Since relevance was already indicated by subjects who had decided to use a specific record, it is assumed the resultant score would be appropriately high as well; in fact, this occurred showing that there is not much difference between English searching (M=81.08) and non-English searching (M=70.38) (about a 15% increase).

		N	Mean	S. D
English searching		75	81.07	20.653
Non-English searching		79	70.38	26.672
Total		154	75.58	24.448
Model	Fixed Effects Random Effects			23.932
	Handom Enoolo			

Table 51. Descriptives of Relevance of intention to use by English vs. non-English search

The ANOVA test results are significant. The test results show the relevance score to be statistically different for English vs. non-English searching (F(1, 152) = 7.67, p=.006). Overall, the hypothesis is supported but the effect size is low (eta sq = 4.8% of the relevance variance explained by English vs. non-English searching) indicating a significant but possibly minor explanation.

Sum of Squares Mean Square F df Sig. **Between Groups** 4394.128 4394.128 1 7.672 .006 Within Groups 87053.274 152 572.719 Total 91447.403 153

Table 52. ANOVA test for Relevance of intention to use by English vs. non-English search

5.5 Research question 5:

RQ 5. What is an explanatory model for users searching for non-English information? How does this model explain influences of user characteristics, language, topic, and task in their query formulation and relevance evaluations of bibliographic information?

An explanatory model was made based on data analysis results previously reported and several data analyses in this section. Before revealing the model, additional data analyses are being done to specify the relationships between variables and the ANOVA factors.

5.5.1. Additional data analyses to assess key variable relationships

A. Satisfaction with the search

There is a cumulative influence of prior variables in the search process which can account for variability in IR system assessment when searching for non-English information. Subjects' satisfaction with the search system can be predicted by the following regression model (N=199). Note that the model explains the equivalent of 87% of the satisfaction with the system variance (F(5,187) = 249.63, p < .0001, R Square = .87). The equation for this model is:

Satisfaction with the search = .537 Efficiency of the system

+ .285 Relevance of intention to use

+ .193 Record understand level

- .079 Difficulty index using the system

- .055 search time

Table 53. Regression model of satisfaction with the search

Summary of satisfaction with the search

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
5	.933(e)	.870	.866	13.846

e Predictors: (Constant), Efficiency of the system, Relevance of intention to use % [0-100%], Record understand level % [0-100%], Difficulty index using the system, Search time (minutes)

ANOVA(f) for Regression model of satisfaction with the search

Model		Sum of Squares	df	Mean Square	F	Sig.
5	Regression	239286.143	5	47857.229	249.633	.000(e)
	Residual	35849.857	187	191.710		
	Total	275136.000	192			

e Predictors: (Constant), Efficiency of the system, Relevance of intention to use % [0-100%], Record understand level % [0-100%], Difficulty index using the system, Search time (minutes) f Dependent Variable: Satisfaction of the search result % [0-100%]

Mode		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		В	Std. Error	Beta		
5	(Constant)	10.190	3.702		2.753	.006
	Efficiency of the system	.524	.041	.537	12.780	.000
	Relevance of intention to use % [0-100%]	.281	.036	.285	7.755	.000
	Record understand level % [0-100%]	.183	.033	.193	5.524	.000
	Difficulty index using the system	-2.198	.882	079	-2.491	.014
	Search time (minutes)	479	.238	055	-2.013	.046

Coefficients(a) for Regression model of satisfaction with the search

a Dependent Variable: Satisfaction of the search result % [0-100%]

	Descrip	otive Statisti	cs for relate	d variables	for the	regression	model
--	---------	----------------	---------------	-------------	---------	------------	-------

	Mean	S. D	Ν
Satisfaction of the search result	39.00	37.855	193
Difficulty index using the system	2.73	1.362	193
Efficiency of the system	39.76	38.814	193
Record understand level % [0-100%]	48.38	39.947	193
Search time (minutes)	5.71	4.311	193

The above analysis yields an interesting model. Users' satisfaction with the search results can be interpreted as one where the perceived efficiency of the IR system overwhelmingly determines the satisfaction level. Note that the standardized coefficient for the efficiency variable ($\beta = .537$) is almost twice the weight of the next contributing variable. This can mean that users are fully aware of system efficiency and view it in terms of how the system serves their needs. In this case, the systems fell short of the user's needs and results in diminished satisfaction scores. The mean satisfaction score on a 100 point scale only achieved an average of 39 points.

B. "Gives up" variable to stop the search

A logistic regression model revealed that six search behavior variables explained the equivalent of 57% of the variability in the categorical decision to terminate a search:

- Topic familiarity;
- Language;
- Average number of terms used;
- The number of query changes;
- Individuals' assessment of search difficulty;
- Time to conduct search.

The following table shows that Nagelkerke R Square (N=288) = .574 and this model correctly predicted 82% of the decisions to give up a search (correctly predicted for **no** give-up: 87.7 %; and, for **yes give-up**: 71.4%). Future research might consider focusing on this dependent variable (where continuing is coded as a 1 and ending the search as a 0). Quitting a search could be considered as an indicator of user frustration, confusion, or fatigue. A key variable in such a study would be the difficulty index which

is a strong indicator of continuing a search. The predictor level for search termination can be assessed by looking at the odds ratio for each variable, expressed in the table (variables in the equation) as Exp(B). For the difficulty coefficient = 1.455, its odds ratio is Exp(B) = 4.283 and this indicates it is the strongest contributor to the decision to quit the search. Importantly, the language variable provides the next strongest indicator with its odds ratio at 2.053. The only other variable exerting influence in this model is the number of query changes with Exp(B) = 1.207. Future studies on search termination in a multilanguage IR environment may begin with these three indicators and add other promising variables to increase the prediction level and the Nagelkerke R Square effect size.

Table 54. A regression model for a give-up search Variables in the Equation for a give-up regression model

		В	S.E.	Wald	df	Sig.	Exp(B)
Step	Topic familiarity	018	.007	7.073	1	.008	.983
	Language	.719	.258	7.776	1	.005	2.053
	Numbers of terms	942	.266	12.499	1	.000	.390
	Query change	.188	.082	5.240	1	.022	1.207
	Difficulty index	1.455	.243	35.862	1	.000	4.283
	Search time	286	.070	16.623	1	.000	.751
	Constant	-1.945	.973	3.996	1	.046	.143

Model Summary for a give-up regression model

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
6	147.524(b)	.419	.574

b Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Classification Table(a) for a give-up regression model

Observed			Predicted				
			Give up Perce		Percentage		
			No	Yes	Correct		
Step 6	Give up	No	107	15		87.7	
		Yes	20	50		71.4	
	Overall Perc	entage				81.8	

a The cut value is .500

5.5.2. An explanatory model of non-English information searching

The arrows in figure 11 indicate relationships which are statistically significantly related to each other within an understanding of necessary or sufficient preconditions for search behavior sequences. The arrows show an approximation of relationships akin to a cause and effect pattern with direction indicators. Again, these arrows all have achieved statistically significant relationships. The text following figure 12 provides a capsule summary of findings depicted in the model.

Figure 11. An explanatory morel of searching for non-English information (all arrows showing)



A. Persons characteristics

Among person characteristics, a person's native language affects each query formulation variable when searching for non-English information and this relationship is statistically significant. Job experience influences the behavior to read bibliographic records which were presented in non-English. In this study of researchers and library professionals, there were different search behaviors (query formulation) and different understanding levels of bibliographic records by the individuals' different professions.

B. Person's experiential knowledge

Topic knowledge and language knowledge both affect users' termination of a search and users' understanding of bibliographic information. Language knowledge for a target task also affects users' search behaviors. Higher language knowledge leads to an increase in the number of different language searches resulting in more query changes and greater search time in a non-English information search.

C. Situation factors (Task, language, topic and system)

Searching with an assigned language results in statistically significantly different query formulation and system choices, bibliographic understanding levels, and satisfaction with the system for English searching compared to non-English searching,. The system's language function and its support exert influence over use and evaluations of the search and the system.

D. Query construction

Difficulty index and number of examined records' in query formulation affect the decision of intent to use the document. Search time and number of terms used affect the subjects' decision to terminate their search. Search time and difficulty index influence subjects' system evaluation.

E. Bibliographic record understanding level

Assigned language knowledge, job experience, topic familiarity, and language knowledge affect subjects' bibliographic understanding level. Note that English and non-English searching have different understanding levels.

F. Intention to use the retrieved information

Relevance assessments, the number of examined records, and the difficulty index affect the decision to use the retrieved record or not when searching for non-English information.

G. Satisfaction with the search

Satisfaction with the search is a function of efficiency of the system, relevance assessment, intention to use retrieved documents, record understanding level, difficulty using the system, and search time.

5.5.3. Non-English information searching: a distilled, brief overview of the derived explanatory model

As seen the figure 5.9 below, this model explains "access and use of non-Roman alphabet language information as a function of:

person characteristics (includes native language, job experience and profession),

- person's experiential knowledge (includes topic, system, and language knowledge),
- situation (includes task, topic, system, language),
- query construction (includes first language choice, search time, number of query changes, number of different language uses, number of terms used, number of examined records, Boolean logic use, and difficulty index),
- IR system preference (includes system choice and users' termination of their search),
- bibliographic record understanding,
- users' satisfaction with the system,
- system efficiency for non-Roman alphabet information, and
- relevance as intention to use retrieved documents.

In all, the model indicates there are influences of person characteristics and knowledge experiences in individuals' selection of retrieval system, query construction, and evaluation and use of retrieved information using non-English language information.



Figure 12. Non-English information searching: a distilled, brief overview of the derived explanatory model

5.6. Summary of major data analyses results

Table 55.	Summarv	of major	data	analyses	results

RQ & Hypotheses	Summary of major data analyses results
1. What are non-Roman	R1.1. System use for non-English information
alphabet users' information	OPAC: local library & nationally famous libraries with Asian collection, such as library
system use and search	catalog systems in Columbia University, Princeton University, Cornell University, Harvard
behaviors?	University, UC library system (MELVL), and OCLC system. Some participants mentioned library
1.1 How do users use current	systems which are located in other countries (Yonsei University library system, KERIS, Washeda
information retrieval systems	University and National Diet Library (NDL)).
searching for non-Roman	WorldCat or RLIN system: an almost bimodal pattern with those systems used either 'not at
alphabet information?	all (30%)' or 'most use' and 'very often use' (34%) from survey participants.
	Online DBs: total of 57.6% (99 out of 175) participants indicated they do not use online DBs
	for non-English searching. Current users are not looking for journal articles written in non-English
	languages because they were unaware which online databases would provide such access;
	additionally, there are not many online databases covering non-English materials. Only 23% of
	participants responded they use online DBs very often: the abstract/journal names known by users
	are "China Academic Journal," "JSTOR," "Magazine Plus," "CiNii," "SpringerLink" and
	"PubMed/Medline."
	Web search engines: web search engines were used the most for searching non-English
	information. Google is the most used. Several respondents indicated that they search regional
	Google, such as google.tw.com which is Google for Taiwanese country. "Baidu" was mentioned
	by several respondents which provides multimedia information and is searchable in vernacular
	languages including CJK. Several regional search engines were mentioned too, such as
	"Naver.com," "Daum.net" etc from South Korea.
1.2 What are the users' needs	P12 A Usors' need for non English information: Research work connect trust translation (went
and expectations when	K1.2. A. USEIS fictu for hon-English mornation. Research, work, cannot fust translation (wall
and expectations when	general information to get information about their home country
information using such IR	B Users' expectation with non-English searching: Cross Language Information Retrieval function

4 9	
systems?	needed, Easy access to full text non-English Information, Interface design, Facilitate multilingual
	query input/search methods, Record display with Translation service assistance
 1.3 What are their language choices and considerations when they construct their queries? H1.3: Users will input (or prefer) to state queries in their native language rather than English. 	H1.3. Hypothesis is not supported. Most subjects' first language input choice was English (90%) from experiment data. From survey data, English was the most chosen in an English environment; target language (52%, n=92) was chosen the most in non-English environment. Native language 23.7% (n=42) and English 22% (n=39) followed this pattern. Findings could be due to use of convenience sample who knew English and other languages.
1.4 How do users' background and their experiential knowledge affect their non-Roman alphabet information system use and search behavior, if at all?	R1.4. From experiment observation: Native CJK subjects' have a better understating of Chinese characters and try them in their queries. They try to input Chinese characters in a query if they are assigned one of the CJK as their target language. The native CJK users use their own reference tool to find a right target word; whereas, English speakers' seems to try in English to retrieve non-English records without any other action such as looking for options in the interface to find translation functions or referring to other tools. Many native English speakers including librarians were not able to find information written in the assigned language. They would need assistance to address access to non-English information.
H1.4: There are statistically significant differences in users' IR system use and search behavior by users' background and their experiential knowledge.	H1.4.1 Search behaviors by their native language background: hypothesis is supported A. MANOVA test: the following search behaviors are statistically different by different language backgrounds for subjects: the number of different language used: $F(3, 195) = 5.81, p < .001$, the average number of search terms: $F(3, 195) = 9.47, p < .001$, the number of query changed: $F(3, 195) = 6.39, p < .001$, search time: $F(3, 195) = 3.56, p < .05$ B. The Chi-Square test: Boolean logic was used differently by different language groups (Pearson Chi-Square (3, N=199) = 16.46, $p = .001$). 60% of subjects used Boolean logic when searching for non-English information. Compared to other language groups, native English subjects used Boolean logic less than non-native English subjects. Advanced search option was used differently by different language groups (Pearson Chi-

Square (3, N=199) = 8.384, p < .05). 74% of the searches used the advanced search options when searching for non-English information. H1.4: There are H1.4.2. Search behavior by researcher vs. library professionals - this hypothesis is supported statistically significant A. One way ANOVA: the following three variables show statistically significant differences for differences in users' IR system the two groups: use and search behavior by Number of examined records: F(1, 197) = 8.905, p < .005users' background and their Difficulty index using the system: F(1, 197) = 8.26, p < .005experiential knowledge. Search time: F(1, 197) = 4.30, p < .05Library professionals examined more records when they were searching, and accordingly the search time is longer than for researchers. Researchers expressed more difficulties than library professionals when they search for non-English information. B. Chi-Square result: Advanced option use shows a significant difference by two groups (Pearson Chi-Square F(1, N=199) = 7.36, p = .007). Librarian groups used more advanced options. H1.4.3. Search behavior by experiential knowledge: hypothesis is supported MANOVA test: search behavior variables including factorial search behavior variables (first language choice and advanced option use) by three experiential knowledge variables (topic, system and language). The number of language changes, difficulty index, examined records and first language choice were shown to be significantly different by the three experiential knowledge levels. In the follow up test, topic familiarity and system knowledge do not show overall significant results. The following variables have statistically significantly differences by assigned language knowledge with a notable 40% effect size when searches occurred for non-English information. The independent measure accounting for these findings is knowledge level for the search topic and task. Number of language changes: F(1, 186) = 14, p = .001Examined record: F(1, 186) = 7.9, p = .006Difficulty index: F(1, 186) = 15.81, p = .001

 1.5 Are user behaviors consistent when accessing English and non-English information via online DBs, OPACs and web search engine? H1.5: There are statistically significant differences in users' IR system use and search behavior for non-English information when compared to English language searching. 	H1.5. Differences by English vs. Non-Eng search (CJK) – hypothesis is supported. One way ANOVA test: three search behaviors and one evaluation of the system are statistically significantly different for non-English vs. English searches. There were more languages used and more records examined in non-English searching than English searching. Non-English searching was more difficult than English searching according to the difficulty index. English searching shows much higher satisfaction scores for searches than non-English searching. The number of different language used: $F(1, 286) = 11.53, p = .001$ The number of queries examined: $F(1, 286) = 7.24, p = .008$ Difficulty index using the system: $F(1, 286) = 98.30, p < .001$ System evaluation: $F(1, 286) = 95.64, p < .001$.
 What issues are present when individuals search non-English information via online DBs, OPACs and web search engine? 1.What kinds of issues and limitations exist when searching non-English information? 	 R 2.1: Issues using current IR systems A. Lack of non-English access via indexing terms B. Lack of non-English records in major online databases which index journals C. Lack of English translation of abstract and summary Lack of coherent and understandable access to non-Roman language materials (Romanization issues - Standardization in Romanization in each language, difficult to type/to know Romanized word, and inaccurate data description.)
 2.2.How do those factors affect the access to information contained in such systems? H2.2.1: Users will avoid databases if the system lacks efficient and comprehensive language coverage when they look for non-English information. 	H 2.2.1 The hypothesis is supported Among three test systems, the Google system has the highest efficiency score (M =66.67, SD=27.26) followed by WorldCat (M =46.03, SD =40.62), and then, far behind these, EBSCO Host (M =8.26, SD =20.46). The satisfaction and difficulty index scores also indicate that the EBSCO Host system is the most difficult to use and this is also affirmed by its low satisfaction score. An ANOVA test was conducted comparing the three systems with various variables. In the experiment (N=32), EBSCOhost was avoided more than the other DBs. An overwhelming majority, 69% (n= 22 people) chose Google to search for non-English information, whereas WorldCat and Yahoo were chosen by only 9.4% (3 people each).
--	---
H2.2.2: If the system lacks efficient and comprehensive language coverage, then users' search time will be shorter than with other systems when they search for non-English information.	H 2.2.2 The hypothesis is supported A. An ANOVA test confirmed that the three systems have statistically different search times (F (2, 189) = 3.69, p <.05). The post hoc comparison test of paired means shows that EBSCO Host and WorldCat test are statistically significant different in search time scores and that experiment participants spent little time when using the EBSCO Host system. B. Cross-tab and Chi-Square tests for termination of a search were conducted for the three systems. Many gave up when searching the EBSCO Host system (Give-up: n=50) for non-English information (N=199) compared to other systems (WorldCat: n=16; Google: n=4). Thus, this hypothesis is supported. The Pearson Chi-Square test result shows there is a statistically significant difference by the three systems on the termination of search variable (X^2 (3, N =199) = 71.16, p < .001).
H2.2.3: Users have better knowledge about scholarly academic journals in their areas written in English than journals written in non- English.	H 2.2.3: 112 out of 204 participants provided well known journals' names in their study or work field. Among 112 respondents only 44 people recall or mentioned the non-English journal names in their area of works. Asian studies area has better recall in this area.

3. How do bibliographic records facilitate or hinder the understanding of bibliographic and retrieved information?	R3. A linear regression model revealed that four variables explained the equivalence of 46% of the variability in understanding the bibliographic record as: R (288) = .676, F (4, 227) = 47.66, p < .0001. The model equation can be expressed as follows: Record understanding level = .462 Satisfaction of the search result + .299 Assigned language knowledge + .215 job experience + .191 Topic familiarity
H3.1: There will be statistically significant differences for users' understanding levels when retrieving a full record which includes English translation, Romanized and vernacular language than with other record displays.	H3.1: Hypothesis is not supported: The result shows that people can understand the bibliographic records well when the record has been either described in English only or Romanized, original language (vernacular) with English translation (N=171).
H3.2: There will be statistically significant differences in bibliographic understanding levels by English searching compared to non-English searching	H3.2: Hypothesis is supported: One way ANOVA test result of bibliographic understanding level by English and non-English searches shows there are statistically significantly differences in the subjects' bibliographic understanding level for English searching compared to non-English searching ($F(1, 230) = 34.46, p < .001$). English searching shows a much higher understanding of the record than non-English searching.
H3.3: There will be statistically significant differences in bibliographic understanding level between researchers and library professionals.	H3.3: Hypothesis is supported - The bibliographic record understanding levels vary by researchers vs. library professionals as revealed with a multivariate analysis of variance (MANOVA) (Wilks' Lambda .893, $F(4, 283) = .103$, p<.001 with 11% of variability/effect size explained (Eta Squared = .107)). The descriptive statistics reaffirms that library professionals understood the bibliographic records better than researchers. Researchers have a better understating with English translation only (without original language transcript) than with the full record.

4. What explains non- English information users' relevance certainty judgments (intention to use retrieved information)? How do these intentions compare to those reported for English or monolingual users? H4: There is a statistically significant difference in relevance certainty judgments for searches seeking non-English information compared to searches seeking English information.	 R4: A logistic regression model revealed that three variables explained the equivalence of 90% (Nagelkerke R Square = .896) of the variability in decisions on whether to use the retrieved record or not when searching for non-English information (N=199): Relevance Number of examined record Difficulty index This model correctly predicted 97% of the decisions to "intend to use" the record's document (correctly predicted for no use: 99%; and, for yes: 93%). Other variables, such as language knowledge and system knowledge were expected to influence this decision as well but they were not statistically significant in explaining intention to use the document. The three significant variables are the most powerful indicators affecting users' judgment whether to use the document indexed by the bibliographic record. H4: Hypothesis is supported. The ANOVA test result is significant in explaining relevance certainty judgments. The test result shows the relevance score is statistically different when comparing English vs. non-English searching (<i>F</i> (1, 152) = 7.67, <i>p</i>=.006).
5. What is an explanatory model for users searching for non-English information? How does this model explain influences of user characteristics, language, topic, and task in their query formulation and relevance evaluations of bibliographic information?	An explanatory model was established. This model explains "access and use of non-Roman alphabet language information as functions of: 1) person characteristics (includes native language, job experience and profession), 2) person's experiential knowledge (includes topic, system, and language knowledge), 3) situation (includes task, topic, system, language), 4) query construction (includes first language choice, search time, number of query changes, number of different language uses, number of terms used, number of examined records, Boolean logic use and difficulty index), 5) IR system preference (includes system choice and users' termination of their search), 6) bibliographic record understanding, 7) users' satisfaction with the system search, 8) system efficiency for non-Roman alphabet information, and 9) relevance or intent to use identified documents retrieved.

Chapter 6: Discussion, Implications, and Future Research

6.1. A derived, empirically based explanatory model of non-English information

searching

Figure 13. An explanatory morel of non-English information



The derived, empirically based explanatory model of non-English information searching (Figure) attempts to depict a comprehensive view of how users interact with non-English (including non-Roman alphabet) information systems. This model is complex because user/system interaction is complicated. In fact, the search process can be viewed from other perspectives and can be studied using other methodologies. The model presented above is an attempt to capture what happens when individuals with varying language competencies attempt to search for information on different topics using different IR systems. This model evolved from two parallel studies: an experiment and a survey. A focus here is to account for the variability in users' behaviors when using different languages with different alphabets while seeking information in standard or multilingual data bases. The arrows in the figure indicate there are statistically significantly relationships for the direction between each concept and its variables. In this section each relationship between variables and/or concepts will be reviewed.

6.1.1. Effect of person's characteristics on searching for non-English-information

6.1.1.1. Different search behaviors by four language groups (CJKE)

As results of the data analyses, it was discovered using convenience samples that an individual's native language background affects each query formulation which impacts: search time, number of query changes, number of different language changes, number of term uses, number of examined records, Boolean logic use, advance search option use, and assessment of difficulty when searching for non-English information. These search behaviors were different for each CJKE language background and also different between CJK and English background searches.

Most subjects, including English speakers, claimed that it is very difficult to understand the Romanization description without prior knowledge of the record or special expertise in the original language. The problems were less pronounced for Japanese compared to other CK subjects who were better able to read the Romanization for Japanese materials. This is probably due to the Japanese use of a Romanization scheme when they type in Japanese using word processing. Note that Koreans do not use Romanization to type Korean using word processing and that their computer key boards contain the Korean alphabet.

Most native CJK subjects have a better understating of Chinese characters and try Chinese characters in their queries when they search for CJK information. Since these three language groups share some parts of Chinese characters, when the CJK subjects were assigned a task in CJK, especially when they do not know the target language, many subjects tend to try Chinese characters in their searches. Actually there are very few times when these three languages share the same Chinese word characters, but users might try to assume the meaning of the letters when using a similar shape containing a Chinese logogram representing a word or morpheme character.

The CJK native subjects tried to use their own reference tools to find the right target word when they were assigned a CJK language they did not know. The reference tools they used were mostly an online dictionary or web search engine from their native languages, such as yahoo.co.jp (Japanese), google.co.tw (Taiwanese), and Naver.com (Korean). Many CJK experiment participants also tried those systems to get a better idea before they actually started their search.

It is interesting that most CJK subjects seemed to know how to process their search task even though they did not know the assigned language, whereas, English speakers needed assistance when they were assigned to search for CJK information. Many native English speakers including librarians were not able to find information written in the assigned language.

It is assumed there will be more issues related to such cultural and language

differences that should be addressed when structuring a multilanguage IR and CLIR system for target users. This study offers a model which affirms that it is important to understand differences among different cultures with different languages and different alphabets when designing a user focused IR system.

6.1.1.2. Different search behaviors by different professions

It has been found that there are different search behaviors in query formulation and different understanding levels of non-English bibliographic records by researchers and library professionals. Compared to librarians, researchers took more time searching and indicated greater difficulty in conducting searches. In a way, this could appear obvious but it can also be interpreted that current IR systems are not supportive of searching for non-English materials by an informed public.

Library professionals examined more records when they were searching, and accordingly, their search time was longer than the times recorded by researchers. Researchers expressed more difficulties than library professionals when they searched for non-English information. Boolean logic was used at almost the same level for both researchers and library professionals. In advanced search option use, library professionals used the options more often than researchers.

6.1.1.3. Bibliographic understanding level by profession and job experience

Among person characteristics, two variables — profession and years of job experience — explain variability in users' understanding of bibliographic records for non-English information. Library professionals had a better understanding of non-English bibliographic records than researchers and this might be due to their ability to understand the components of a bibliographic record. Additionally, the more job experience an individual had, the better the understanding of non-English bibliographic information.

6.1.2. Effect of person's experiential knowledge on searching for non-Englishinformation

Topic knowledge and language knowledge both affect of the decision to terminate a search. Topic and language knowledge also explains the users' understanding of bibliographic information when searching for non-English information.

Language knowledge for the target task affects users' query formulation behaviors. The higher the users' language knowledge, the greater the number of different language searches, the more there are query changes, and the greater is the search time when seeking non-English information. It is obvious that there will be differences when searching for information between people who know a target language and those who do not know the language. Yet, should this be the case if the IR system were truly user friendly? A conjecture might be offered here that a good multilanguage or cross language IR system would occur when users' language knowledge does not have much influence on using a specific system. We could ask how the system might adjust or even reconfigure itself to accommodate users' different language knowledge.

6.1.3. Effect of situation factors (Task, language, topic and system) on searching for non-English-information

6.1.3.1. Searching with a different language

It is found that searching with an assigned language results in statistically significantly differences in query formulation, system choice, bibliographic understanding level, and satisfaction with the system. These differences were heightened when comparing English searching with non-English searching. The functionality and design of the system's language functions exerts significant influences on users and their assessments of different IR systems.

Three search behaviors (i.e., the number of different languages used, the number of records examined, and the reported difficulty in using the system) are statistically significantly different for non-English vs. English searches. There were more different languages used and more records examined in non-English searching than in English searching. Given the capabilities of the systems, it was not surprising that non-English searching was consistently reported to be more difficult than English searching. Each CJK language search was reported to be more difficult than English searches. A comment is needed regarding this finding: it implies that even for individuals knowing different languages and alphabets. English is still valued in the search environment—at least by this convenience sample. It was consistent, then, that findings also revealed that English searching showed a much higher satisfaction score and bibliographic understanding level than non-English searching.

6.1.3.2. Searching in three IR systems

From various data analyses, it was found that the EBSCO Host system was regarded by users as the most inefficient system among the three systems when searching for non-English information when compared to WorldCat and Google. When subjects in the experiment were asked to choose a system to search, EBSCOHost was chosen the fewest times. Also, EBSCOHost use resulted in the highest number of search terminations.

As reported previously, users were least likely to use online database searching for non-English journal articles. One reason reported for this is that language supportive systems for non-English journal articles were more likely to be found in web search engines and the WorldCat system. The other reason offered for selecting a source would be the systems' perceived coverage of non-English materials. Among 112 out of 204 participants, only 44 people recalled or mentioned non-English journal titles in their areas of study or expertise. One survey participant asked why there was not better journals access to non-English source materials as "many works are not accessible from major English database." The perceived availability of information becomes a critical component in the information search process and this adds yet additional support to Kuhlthau's four criteria affecting information seeker's search process behaviors: task, time, interest, and availability of the information (Kuhlthau, 1993a, p. 39),

6.1.3.3. Different use behaviors when using the three different systems

A. WorldCat: Most users seemed uncertain when beginning to type their queries to seek non-English information. Most users typed in English and most of the time they did not get good results. For example, subject #27 said "I am so frustrated it seems it has some information but I don't get any results. Is this because I did search in English? How do I know equivalent word for that?" Even CJK native subjects appeared to have difficulties finding information in their own language. These difficulties spanned query formulation to reading and understanding the bibliographic records retrieved. Many CJK subjects tried their own native language and romanized words for their queries but the amount of retrieved information was very small and many records only contained romanized descriptions without their own language description and English translation.

Most subjects, especially those who did not know about WorldCat before, were surprised the system had such a huge collection of non-English materials and that it allowed users to search in their own languages. At the same time, the users stated that they wanted a better index system to access more information in the collection for any or all languages.

B. EBSCOHost: Although it has many good qualities with its inclusion of databases on various subjects, its use was not transparent. First, there were substantial differences in the availability of non-English collections depending on the database that EBSCOHost was accessing. Many of its databases did not seem to have very much non-English information. Only one expert subject knew how to use the advance search option to 'limit' results to a specific language collection.

Although users were able to locate non-English records, many of these items do not provide an English translation of the description which includes neither title nor abstract. Romanized descriptions prevailed and subjects could not decide from this if a particular record was relevant or not. Also users were surprised that the system did not allow one to type their query in languages other than English. The language box on the first screen appeared to users that it would allow them to type in different languages and retrieve in those languages; however, it only allowed them to change the input screen's interface. Most subjects also expressed that it was inconvenient that the query input box only allowed one word in each box.

C. Google: Most subjects had extensive prior experience with Google searching although there was concern about the quality of the information retrieved. Even when they used Google scholar, the users' felt the retrieved information may not be from an authentic information source.

Most subjects including native CJK subjects were surprised that Google has a language tool where they can translate the query and retrieve information in other languages. Also they seemed happy to have the option which translates a website. Native English speakers, including librarians, did not report that they would be able to use the language service but they were amazed at its capabilities. Nonetheless, most subjects reported on the poor quality of the translation provided from Google. Yet, Google became the model for an IR system which starting to approach the functionality of a CLIR system. Google allows for complex entry of queries, it allows for queries to be stated in different languages, it allows for retrieval in different languages, and it has an elementary machine translation capability. Based on the evidence presented by the convenience sample used in this study, other IR systems might look to these functions as they design improvements to their existing systems.

6.14. Language choices and considerations regarding query formulation

In this experiment, most subjects' first language choice was English (90%, n=225) when they searched for non-English information (N=252). It might be assumed

that this language decision was due to this particular experiment where subjects were mostly students and librarians who lived in the U.S. and were using English for various purposes. Also, the experiment was primarily done in an English language environment. Yet, given the composition of the sample of subjects, it was still surprising that only about 10% of the search tasks were tried in either 'their native and target language' or their 'native' or 'target language.'

In the online survey (N=178), English was chosen for their query language by just over half of the respondents: 55.1% (n=98) in the English dominant system environment; whereas, the target language was chosen for their query language also about half the time at 52% (n=92) in the non-English system environment (N=177).

This leads to a tentative conclusion that a user sensitive system would have CLIR functions where individuals can specify language choice for each aspect of the search process: query, database languages queried, and target languages retrieved. As will be seen in the next sections, this recommendation will also extend to the retrieved bibliographic record and to the language alternatives once documents have been identified for retrieval. This will also have implications for translation alternatives.

6.1.5. Bibliographic record understanding level

6.1.5.1. Variables affecting the understanding of the bibliographic record

Four variables were identified as important (i.e., statistically significant) in their influence over the subject's understanding of the non-English bibliographic records: satisfaction with the search result; assigned language knowledge; job experience; and, topic familiarity. Remember that it was previously reported that there were also

significant differences reported for individuals' profession and understanding level when conducting English vs. non-English searching.

6.1.5.2. Reading different bibliographic records

The online survey respondents indicated that the record structures receiving the highest scores were for 'English only description' and 'Romanized with original language (vernacular) and English translation.' Common expectations that survey participants mentioned were:

- Need English translations accompanying detailed information in the record
- Need original scripts and romanized words
- Brief description of title's contents (plot summaries, themes)/ story titles more specific contents information (i.e. chapter by chapter)
- Description needed (e.g., abstract)
- Link to records of other editions of this book available
- Links to other Romanization records for the same author and title

Note that respondents not only indicated a need for assistance with language issues, but that they also wanted to have a brief description about the work in any format (such as abstract, summary or plot, and table of contents).

There were significant differences by researchers and library professionals in their understanding of the bibliographic records for non-English materials from the online survey participants. Library professionals understood the bibliographic records better than researchers, especially records with romanized and vernacular descriptions. Researchers reported a better understating of English translation records (without original language transcript) than full records. The inclusion of these variables in this research was based on the premise that bibliographic records should be as well understood by public users including researchers as it is by information professionals like librarians. In short, the Romanization systems currently in use do create some confusion and this study shows that they need to be augmented by enhanced records.

6.1.6. Non-English information users' relevance certainty judgments (intention to use retrieved information)

A logistic regression model revealed that three variables (i.e., relevance score, number of examined records, and difficulty index) explained the equivalence of 90% of the variability (Nagelkerke R Square = .896) in decisions as to whether to use the retrieved document or not when searching for non-English information (N=199). These three variables represent powerful factors affecting users' judgments on whether to use the bibliographic record and the document it indexes. The relevance score by English searching and non-English is statistically different for these two types of searches (English searching has higher scores than non-English searching). It had been proposed that additional variables also affect users' relevance intentions to use a specific bibliographic record to examine a document and that such variables include language knowledge and system knowledge. Yet, these two variables did not achieve statistically significant results when testing that part of the model.

6.1.7. Query construction

In this study various search behaviors were tested: search time, difficulty index, advanced search option use, Boolean logic use, the number of query changes, the number of language changes, the average number of terms used, and the number of examined records. Individuals' query formulation when searching for non-English information is explained by person characteristics and the search environment. The specific variables explaining query formulation include: the individual's native language, the person's knowledge of other languages, and the system and its interface. It is noteworthy that these independent variables could be cast as language related factors which indicate how carefully this needs to be considered in every step of the system design.

6.1.8. IR system preference

Intention to use the retrieved information is explained by the difficulty index and by the number of records examined. The decision to terminate a search is explained by person, knowledge, and language characteristics. Person's characteristics (job experience and profession), person's experiential knowledge (topic & language knowledge) and different language situations affect the subjects' decision to end their search. The users' decision to terminate searches was hypothesized to be directly tied to the system's support of language functions for non-English searching. In this study, many subjects in the experiment had to stop their search when using EBSCOHost because they reported it had an inefficient language function.

6.1.9. Satisfaction with the search

Satisfaction with the search when seeking non-English information was explained by several variables: perceived system efficiency, relevance rating indicating the users' intention to use the retrieved information, the retrieved information assessment, record understanding level, difficulty index using the system, the time for conducting the search, and knowledge of different languages. These factors and variables might be considered as core criteria for non-English information searching using multilanguage and cross language IR systems.

This model was made based on analyses of statistical data collected for the two parallel studies presented here: the experiment and the survey. The model's explanation can be stated as follows : "access and use of non-Roman alphabet language information is a function of 1) person characteristics (includes native language, job experience and profession), 2) person's experiential knowledge (includes topic, system, and language knowledge), 3) situation (includes task, topic, system, language), 4) query construction (includes first language choice, search time, number of query changes, number of different language uses, number of terms used, number of examined records, Boolean logic use and difficulty index), 5) IR system preference (includes system choice and users' decision to terminate their search), 6) understanding of the bibliographic record, 7) users' satisfaction with the system search, 8) system efficiency for non-Roman alphabet information, and 9) relevance intention to use retrieved information.

This study confirms what may appear obvious but has not been found to be investigated in such user/system studies: <u>that language background (including native</u> <u>language and language knowledge) and searches in different languages account for</u> <u>important variance in explaining non-English information seeking</u>. In this user study, CJK and English languages were tested and these language factors affect almost every aspect of search behavior and system evaluation.

6.2. Current IR system use

6.2.1. System use when searching for non-Roman alphabet information

One particular finding emerged from the survey regarding users: they have low expectations and actual low use rates when seeking online journals which contain non-English information. Participants mentioned that it was their perception that most current major online databases for journal information do not cover many non-English language materials and that the IR systems providing access to these journals do not have language supportive retrieval functions. Some specific database systems were mentioned many times in this survey group which includes: "China Academic Journal (CAJ)," "JSTOR," "Magazine Plus," "CiNii," "SpringerLink" and "PubMed/Medline." The CAJ is a database providing full-text electronic versions of Chinese academic periodicals in various subject areas. This database also provides Chinese full text with English title and abstract. It is an important fact that current non-English users need not only full-text with the original language but also English title and abstract.

The other noticeable difference between general English searching and non-English searching is that half of the participants mentioned they use particular OPAC systems well known for their Asian collections, such as the library catalog systems at Columbia University, Princeton University, Cornell University, Harvard University, UC library system (MELVL), and OCLC system. Some participants even mentioned library systems located in other countries because the current OPAC systems available to them do not support non-Roman alphabet searching. These individuals reported that they use these external OPAC systems to gain better access to collections of non-English information (and this mostly concerns CJK information). Among current IR systems, web search engines are the most used systems when people search for non-English information. Among these, Google is the most used search engine. Participants mentioned that Google is the most convenient and that it is similar to OPAC use. Several respondents use regional search engines in their native languages, including Google, such as google.tw.com, Amazon.co.jp, Baidu, Naver, and Daum etc. For non-native English speakers, obtaining information in their native languages was reported as important because then users can understand the information retrieved better than with an English-only system. Nonetheless, when confusion arose from such conventions as Romanization, respondents preferred that parallel information be available in English.

Not many respondents seem to know that Google has language search tools which provide a version of a CLIR function. A number of survey participants noted that China Academic Journals (CAJ) allows users to type queries in English and return full-texts in Chinese. It also provides English abstracts too. Meanwhile, many respondents seem to be confused about what a CLIR function means since people mentioned WorldCat, Yahoo, Baidu, and Wikipidia as systems which contain CLIR functions. Only two systems provide this function: Baidu can read and retrieve pinyin words which is one of the Chinese Romanization rule systems, and WorldCat/RLIN also provides access for Chinese. Many University library systems such as those at UCLA, Duke, University of Chicago, etc, were also mentioned as systems which provide CLIR functions. These systems allow limited vernacular retrieval functions since they have well organized bibliographic records, vernacular, Romanized, and English records. Thus, users can search either vernacular languages, Romanized words, or English if that word is designated in the record. Survey respondents were unsure of what a fully functional CLIR system would include and, yet, it is the subject participants in this sample who might be expected to have a better knowledge of the capabilities of such systems. It can be concluded that the lack of many CLIR systems leaves knowledgeable users without a mental model of what such a system might provide.

6.2.2. Users' needs and expectations when accessing non-English information

Half of the survey respondents, 97 subjects (out of 192, 50.5%), reported that they need non-English information for their research. This need rose to two-thirds of subjects when various personal interests were taken into account with 32 additional subjects (16.7%) also indicating that they needed access to non-English information. One subject mentioned that one particular country has available advanced information about video games and implied that this information could be made available to a larger audience if IR system tools made it possible.

Users noted their expectations accessing and using IR systems to search for non-English information. Many participants mentioned they want a system that accepts queries written in any language and retrieves results in any other language. Participants were specific in articulating their expectations from retrieval function to interface design. In all, the expectations can be summarized into two major concerns: 1) one is that respondents want to have a system that can search and retrieve information on a cross language basis; and, 2) the other is that they want to have informative bibliographic records which can be understandable to people who may not know the target language. As of today, such an understandable description could be provided using an English translation of the bibliographic record which also includes an English abstract.

For the interface expectation, many participants mentioned they want to have a system interface similar to the one Google provides. In other words, they do not want to have complicated interfaces but, instead, ones which seem to require little knowledge about the use of the interface. Many survey and experiment participants expressed they liked Google because it is easy to use and it does not seem to require new knowledge to use the system when compared to other IR systems which can have complex advanced search protocols. With the exception of a few experts, the subjects in the experiment took time to familiarize themselves with the systems. Most participants typed in more than two queries in a single dialog box which are not allowed in WorldCat and EBSCOHost system and these individuals commented that such an interface is not convenient to use.

Importantly, this study reaffirms findings from other studies. One premise underlying this study was that users' information seeking behaviors are partly explained by their role, language, and experience. Resnik & Oard (1999) had also reported that the CLIR users' background was crucial in understanding the design requirements of a CLIR system and this study confirmed that finding. Also, this study explored the interface available to users and, again, findings emerged similar to those found by Resnik & Oard (1999). Finally, Resnik & Oard (1999) proposed that clear display of retrieval results would facilitate search effectiveness and, again, this study confirmed that proposition.

This study also reinforced but did not completely confirm the findings obtained by Jansen, Spink, and Saracevic (2000) who reported that 67% of Excite.com search engine users submit their queries using a single word. For the most part, users of this study tried to use multiple words in queries and debriefing indicated that experience with Google reinforced this behavior. Even so, this study, like that of Spink & Jansen (2004), found that users do get confused by retrieval results. It is suggested here that behaviors of users of search engines, bibliographic databases, and other IR systems may have a common set of expectations and behaviors common to all information seeking. When a strong variable such as language enters into the task and search requirement, then it can have a mitigating effect on how users begin seaches, conduct them, and terminate searches.

6.2.3. Issues when accessing non- English information

A. Lack of non-English access via indexing terms

Participants expressed there are difficulties and limitations to type and search in their target languages. This issue confirms again the need for cross language information retrieval functions in a user sensitive system.

B. Lack of non-English records in major online databases which index journals

Many participants complained about the lack of non-English material coverage in current online database systems.

C. Lack of English translation for abstract or summary

Users often encountered a bibliographic record without abstract or summary and noted that it is hard to judge whether that record would be relevant for them based on reading only the title and, possibly, index terms or descriptors. If the bibliographic record were written using a Romanized (or transliterated) record, it then represented a set of symbols rather than letters conveying information for those who do not know the language.

D. Lack of coherent and understandable access to non-Roman language materials

Romanization was introduced in order to transcribe non-Roman alphabet letters to Roman letters. Romanization has inherent problems based on it very structure since many non-Roman letters do not have exact equivalent Roman letters obviating the isomorphic linking of a Romanized morpheme to the original language character. Romanization issues include its standardization in each language, difficulty in typing or even knowing the Romanized word, and inaccuracies in data description which are especially confusing when there is no parallel information in a language known by the user.

6.3. Reliability, Validity, and Limitations of this study

Both validity and reliability are important factors which can limit the generalizability of a research study. The current investigation included an experiment and a survey and, from these, consistency in responses and consistency in findings did emerge and this would speak to one aspect of the study's reliability. Both research methods produced similar results which act to reinforce findings. Nonetheless, it is recognized that the researcher created the original model for the studies and that the ensuing instruments to record responses were similar in content and even wording. In all, even though reliability may appear reasonable, this issue takes on lesser concern than validity. And, validity in these studies is seriously challenged on several fronts. Before presenting these limitations, it is crucial to remember that the purpose of this study was to augment user/IR models with the inclusion of user experiences when seeking non-English information with non-Roman alphabets. Hence, this study does have notable limitations but it also attempts to provide a foundation for further study in depicting how users could interact with CLIR systems.

If a study can be generalized to the real world, then the study might be said to possess a measure of validity (Littlejohn, 1992, p. 36). Tague-Sutcliffe (1992) provides an example of such validity by asking "Can results obtained with student subjects be replicated in a corporate information center?" (p. 467). This can be considered as an indicator of external validity. Although the current study used limited language samples — Chinese, Japanese, Korean and English languages — the instruments, process, and the results of this study can be applied to other non-Roman language population settings. The methodology of this study including data gathering and data analysis can be applied to other studies where the search behaviors of humans is explored as they interact with IR systems. Also the instruments for the survey questionnaires and the experiment might be applicable to future studies.

A serious limitation of this initial study in this area is its use of a convenience and purposive sampling of individuals whose native languages are Chinese, Japanese, Korean and English. Since the sample was not randomly selected, it severely decreases the generalizations available from such a study and it limits validity beyond the sample. However, the purposive sampling did allow a very specific content to be addressed which might not be easily found from the general public. The study's focus on this specific context (Non-English users) and it provides, it is hoped, interesting and informative insights on the complexity of user/system interactions in a multi-language and multialphabet environment.

Krathwohl (1998) states that in order to have internal validity in research, it has to prove its "conceptual evidence" (rationale for relationships and translation from concept to operational definitions), "empirical evidence" (expected results) and a credible results (p. 139). Overall, this study attempted to test a complex model with many concepts and derived variables which still had applicability to a real life situation. The environment created to test this was as realistic as feasible at this stage of this research. Hence, the internal validity of the study was threatened by extraneous variability introduced by different systems, different tasks/topics being searched, and different users in their role and their language knowledge. Future studies might need to control for such variability and only study one group using one system for one topic. This would be valuable but it would avoid an underlying premise of the current study: to assess real systems with individuals who represent real users. Another threat to internal validity is the large number of variables studied here where the overall sample sizes cannot support extensive breakouts of the ensuing data. Future research might be able to capitalize on the results obtained here by using the most promising variables reported here in those future studies and then maximize effect sizes for findings.

Both the online survey and the experiment instruments had several pre-tests including running preliminary statistical tests in order to assess the variables and the analytic results according to each research question and hypothesis. This data collection and other instruments were revised several times by pre-testing them and the procedures with experts from Asian studies librarians and students who search non-English information on a daily basis. The statistical analyses used in this study are fairly standard and reasonable ones that have been used in other IR studies. This study also tried to employ multiple methods in order to raise its validity: online questionnaires and an experiment with observation and a short exit interview. Future research might map the data to the model by giving more attention to non-standard statistical tests to include development of nonlinear models.

The focus here, on Chinese, Japanese and Korean, also represents different cultures as well as different alphabets and languages. Future research might find it desirable to compare other languages which might generalize the model derived in this study.

6.4. Additional Implications for future research

Future continuation of this research can take the following directions: (1) providing more in-depth research on the three countries and three languages using a more representative sample; (2) expanding the countries surveyed, the languages used, and the number of individuals contacted in each country; (3) comparing different types of multilanguage systems, such as those used by Amazon.com and/or online catalog systems, by different language backgrounds; (4) investigating users' query and search terms. Of special note will be the socio-cognitive and cultural perspectives of the individuals from each country which was not explored in depth in this study.

Another future area for research exploration would be the process of the full and unbridled sharing of bibliographic information across borders and across languages. Within this would be some exploration of the cooperative work now being done, with much of it under the leadership of the Online Computer Library Center (OCLC), which currently directs the WorldCat effort. The current report by the Library of Congress provides promise that major issues in the sharing of bibliographic information are receiving attention. A cohort of LIS researchers is needed to help this advance with appropriate attention to international perspectives: culture, language, access, and record content.

A comprehensive user analysis would be required if a new and efficient CLIR system were to be created. An advanced, ideal information retrieval system for non-English information including multilingual users will require understanding: each different user's language abilities, the underlying cultural implications of the query, and each culture and language's characteristics as well as knowledge of documents (representations) and system. To offer a reasonable interface, when users process their information search, it should offer various options to approach and obtain the target queries or information. For example, this suggests that broader terms or common words are used within a particular language domain and that the system has the ability to offer synonyms with easy cross-referencing access. In short, the user responsive system of the future may need to reconfigure itself based on user requirements.

6.5. Conclusion

This study attempts to depict users' current multilanguage IR system search and use behaviors. It addressed such issues as transliteration versus translation, the use of English as a common touchstone across diverse languages and alphabets, and the difficulties in formulating queries and evaluating results in languages other than those known by the user. Through this study, it is recognized there needs to be better transmission of cultural understandings across countries. Different native languages often engender different perspectives and these express themselves in unstated needs for those using current multilanguage IR systems. It was also found that there are statistically significant differences between English and non-English searching of bibliographic information and IR systems' use.

The experimental results uncover confusion about multilanguage formats and about Romanized, transliterated information. The experiment confirmed the need to use English as a Rosetta Stone to locate records and interpret bibliographic and summary data. This was reinforced in the survey where respondents reported that they could not understand Romanized, transliterated data, even from their own language. It was further confirmed that the widespread use of English in each country's journal literature now makes it appropriate to translate, not transliterate, information to achieve wider and fuller access to and understanding of scholarly information. Importantly, non-English information seeking behaviors are specific, even unique, from findings reported in samelanguage user IR models.

There are certain differences in pursuing information by users with different language backgrounds and currently most systems are not ready to support users seeking information across languages. It is recognized that machine translation functions are not yet stable but this is an experiment in the making and it is ripe for future user/CLIR research. It might also be helpful if systems provided for query expansion functions such as correct word suggestion, synonym, thesauri or distinguishing homophonic words and if it did this in as simple and direct way as possible. Most users want to have an abstract or summarization of a document or book in their language — as well as in English. Thus, the respondents here preferred a system whose bibliographic description included three features: original language, Romanization and English.

User expectations became a focal point several times in this research and this aspect merits closer examination in future research to assess the difference in how users

assess query construction and retrieved information for imposed topics compared to selfgenerated topics. Users differed when searching similar information and future studies might use a mental models approach to map these patterns within a CLIR environment. One variable used in this study—the decision to terminate a search—could be folded into a more general expectations model bounded by the way users visualize the information retrieval process when building bridges across languages and alphabets.

No single study was located which provided a comprehensive view of how users interact with non-English (including non-Roman alphabet) information systems. Nor was a study located which focused on explaining the variability of users' behaviors when using different languages with different alphabets while seeking information in standard or multilingual data bases. It is hoped that this investigation contributes to the development of such a user model. Such research would need to be sensitive to users who are from different social and cultural backgrounds. Searching for non-Roman alphabet information in current IR systems, especially in an English dominant environment, requires defining situations where users deal with more than one language, different tasks, and different systems. The design of a user sensitive system would need to attend to each language's characteristics, representations, and specific needs for a particular intelligent interface.

It became apparent during this study that a prototypical IR system is emerging which offers convenience and transparency to end users: Google. Google facilitates query construction and retrieval in different languages and it offers a rudimentary machine translation service which can augment or, in some cases, replace Romanized transliteration. All these factors are essential or necessary prerequisites for structuring multilanguage IR system components and features. Also, it is hoped that the system could produce more efficient representations of responses to users' queries within the framework of a culturally sensitive and helpful interface. Finally, these goals might be better achieved by recognizing that theories differ by how they account for variability. This study draws on the concepts and variables which were productive in past research while offering new variables for consideration. As such, it is hoped that the survey and experiment presented in this study will help calibrate needs linked to system capabilities in a multilanguage IR environment.

We live in a world that is increasingly interconnected. Globalization is becoming more important in all parts of contemporary society. Easier access to information from across the world occurs through the use of the Internet, but there are still barriers to access that valuable information. Especially for multilingual users, a more efficient and helpful interface is a key component in improving the searching process because it can be expected that most users do not have a perfect knowledge of languages different than the user's native language. The results from this study might be to shed light on those system designs which can be appropriate to real users' needs and difficulties. It is hoped that this study might aid in understanding CLIR users as well as identify system characteristics important in cross language searches.

APPENDICES

Appendix 1: Experiment

- 1.1. Consent Form
- 1.2. Instruction form for experiment
- 1.3. Example of task sheet
- 1.4. Post task questionnaires
- 1.5. Exit interview questions
- 1.6. Exit questionnaires
- 1.7. Note for each task

Appendix 2: Survey questionnaires

- 2.1. Email advertisement for survey
- 2.2. Survey questionnaire

Appendix 1: Experiment

1.1. Consent Form

Experiment on "Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems"

The School of Communication, Information, and Library Studies at Rutgers University supports the practice of protection for human subjects participating in research. The following information is provided for you to decide whether you wish to participate in the present study. You should be aware that even if you agree to participate, you are free to withdraw at any time without penalty. However, by completing the study's questionnaire, you acknowledge the receipt of this consent form and are stating your willingness to take part in this research study.

You are invited to participate in a research study that is being conducted by YooJin Ha, the Principal Investigator of this study, who is a doctoral student in the School of Communication, Information, and Library Studies at Rutgers University. The purpose of this experiment is to identify what kinds of difficulties individuals experience when accessing and using non-English information from current information retrieval systems.

You will be asked to search assigned topics using different information retrieval systems. After your search, you will be asked your opinion about the search experience and the system. There are no correct answers for the search results. Your responses will merely reflect your experience with the information retrieval systems. Completing the experiment should take approximately one hour. You will be observed by the researcher during the online search to record your experiences with the online information retrieval systems. Following the search, the researcher will ask you several questions during a short interview.

The results of this experiment will be reported anonymously. Anonymous means that I will record no information about you that could identify you. This means that I will not record your name, address, phone number, date of birth, etc. I will keep any information about you confidential by limiting individuals' access to the research data and keeping it in a secure location. The research team and the Institutional Review Board at Rutgers University are the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at a professional conference, only group results will be provided, unless you have agreed otherwise.

Although no discomfort is anticipated while participating in this study, you are free to stop your participation by not completing the questionnaire if any of the questions make you feel uncomfortable at any time.

If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University at:

Rutgers University Institutional Review Board for the Protection of Human Subjects Office of Research and Sponsored Programs 3 Rutgers Plaza New Brunswick, NJ 08901-8559 Tel: 732-932-0150 ext. 2104 Email: humansubjects@orsp.rutgers.edu

If you would like additional information regarding this study, including results, before or after its completion, please feel free to contact YooJin Ha by e-mail.

You will be given a copy of this consent form for your records. Sign below if you agree to participate in this research study:

I, _____, have read and understood this description of the study [please print] and agree to participate in the study.

With my signature I affirm that I am at least 18 years of age and have received a copy of the Consent Form to keep.

Principal Investigator, YooJin Ha

Signature_____

Date _____

Sincerely,

YooJin Ha Principal Investigator School of Communication Information and Library Studies Rutgers University 4, Huntington Street, New Brunswick, NJ, 08901, U.S.A. yha@scils.rutgers.edu

1.2. Instruction form for experiment

/ /2007

YooJin Ha

Experiment on Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems

Task instruction

You will be asked to conduct a total of nine separate searches. For each task you may be assigned a system and a target language. You may also be assigned a topic and a format for a particular task.

* System

The following systems will be assigned for your search.

- → WorldCat
- → Google Language Tools
- → EBSCO Host database

Please note that the **WorldCat** system is a super library catalog which provides information about millions of books. It also contains some information about other formats such as periodicals and media. WorldCat provides multilingual access to information in different languages allowing users to search and retrieve in a specified language.

Google provides access to many types of information and this study focuses on its recent cross-language capability, **Google Language Tools**, allowing users to search in one language and retrieve in a number of specified languages.

EBSCOhost states that it "offers a variety of full text databases and popular databases from leading information providers." It provides a multilingual interface so that you can change the language of the INTERFACE by clicking on Language in the search bar.

You will be asked to select a system for you to use when searching for information.

* Language

For purposes of this study, you will be asked to find information written in three different languages for each topic: your native language, English, and a non-English/non-Roman alphabet language. The last language will be assigned to you during your search. If your native language is English then you may be assigned to search two non-English/non-Roman alphabet languages.

* Format

You may be asked to search for results in two formats: journal article or book (monograph). Some tasks will ask <u>you</u> to select any format that you wish to search. Usually when you search WorldCat, you will be asked to find a book, whereas when searching at EBSCOhost, you will be asked to find a journal article.

* Procedure

After each search, you will be asked to assess your search experience and to comment on the information retrieval systems you used. During and after each task, there might be separate questions to be addressed by the investigator if it is needed. After all nine searches are conducted, you will be asked to fill it out a final questionnaire and then participate in a short exit interview.

Please know that there are no right answers or wrong answers. The objective of this research is to explore how people search for certain types of information.

Thank you for your participation! YooJin Ha

1.3. Example of task sheet

Task 1

Topic: Fluoride and Health

Please assume that you need information about the effect of fluoride on health. Fluoride is added to water, toothpaste and other products and research indicates it has an effect on health.

It is important that reputable and authoritative information be retrieved to address this issue. One report indicated that the best information in this area was contained in Korean research. Be sure to get high quality information on this topic.

System: WorldCat

Language: your native language

After your search, please answer the following question.
/ /2007

)

YooJin Ha

Experiment on Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems

- Please indicate your familiarity with the topic you searched. [If topic is very familiar, you might indicate 90 or 100% familiarity; if you are totally unfamiliar with the topic, then indicate 0% familiar.]
 ______% familiar
- How understandable was the information on the retrieved bibliographic records or presentation of the document? [when responding, use 100% if it was fully understandable and a lower percentage score if only part of it could be understood]
 _____% understandable

3. Was there any description in the bibliographic record or presentation of the document that you could not understand even if you know the target language? Please neglect if you did not understand professional acronym or jargon.

Yes No (If yes, please give examples:	
4. How satisfied are you with the retrieval results? [0-100%]	% satisfied
5. Do you think you are going to use any of the information you retriev Yes No	ved?
\rightarrow If yes, please provide your assessment: % relev	vant

(If no, please provide the reason: ______ % relevant

6. How efficient was the system when searching for Non-English information?
 _____ NA: if you did not search for information written in different language
 _____ % efficient

7. Please describe in detail any difficulties you encountered in constructing the query.

8. Please describe in detail any difficulties you encountered using this system when searching for Non-English information even of it is your native language.

_____NA: if you searched information written in English.

1.5. Exit interview questions

/ /2007 YooJin Ha

Experiment on Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems

- 1. Could you tell me how you feel when you use the systems in this study?
- 2. Could you describe differences among those systems?
- 3. Could you tell me if you feel any inconvenience, confusion, or experience any problems when using the systems?
- 4. Are the bibliographic descriptions easy to read for you? What is your highest level of concern when searching for non-English information?
- 5. What is your highest level of concern when interpreting non-English information?

1.6. Exit questionnaires

Experiment on Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems

1. Please indicate your system familiarity [give approximate number of uses of each system].

1) How often have you used WorldCat system in past year from today?

_____ uses of system [if no uses, then fill in 0 uses]

- 2) How often have you used EBSCOhost database system in past year from today? _______ uses of system [if no uses, then fill in 0 uses]
- 3) How often have you used Google search engine in past year from today? _______ uses of system [if no uses, then fill in 0 uses]
- 4) How often have you used the online database system that you chose and used today's search in past year from today?

Online database name:_____

_____ uses of system [if no uses, then fill in 0 uses] Online database name:_____ ____ uses of system [if no uses, then fill in 0 uses] Online database name:_____ ____ uses of system [if no uses, then fill in 0 uses]

2. Could you please indicate under what circumstances you might need to access or use information written in NON-ENGLISH languages?

3. Could you please indicate when was the last time you needed to access information in NON-ENGLISH languages? _____ days ago from today (if never or more than a year ago, please state so).

4. Could you please indicate when was the last time you needed to access information in languages that you cannot read? What was the language? Why did you need that? (if never or more than a year ago, please state so)

_____ days ago from today

 5. Please indicate the specific system name which you use THE MOST OFTEN when you search for NON-ENGLISH information. Please also indicate why you use that particular system the most?

The system's name:

The reason you use this system:

6. If you could create a new system to search for non-English information, what might you create? What kinds of functions and services might be included to help your searching and access for NON-ENGLISH information?

7. What is your native language?

- 8. Where were you born? _____
- 9. If you were not born in the US, how long have you lived in the US: _____ years
- 10. Except for your mother language, please indicate your knowledge of other languages. Using a 100 % scale, indicate your reading comprehension level, speaking level, and writing level for each language you list. A score of 100% indicates complete fluency in that language.

Other Languages: [please fill in]	<u>Reading</u>	<u>Speaking</u>	<u>Writing</u>
	%	%	%
	%	%	%
	%	%	%
	%	%	%

11. Please indicate your academic degrees. Please specify all the college/university degrees if you have (or date by which you expect to be awarded the degree)

Degree	Major	Year

12. Please indicate your current position? Please provide appropriate specific information. (i.e., professor, Asian studies in Rutgers University; librarian, Asian language cataloger, Rutgers University; doctoral student, communication, Rutgers University)

13. If you are/were librarian or professor, please provide total years of your work

experience: _____ years

14. Overall, for how many years have you been doing online searching? _____ years.

1.7. Note for each task

/ /2007 YooJin Ha

Experiment on Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems

Subject ____ Task1 1. Query note

2. Number of queries changed: _____

3. Number of examined records:

4. Search time: _____

5. Comment:

Task2 1. Query note

2. Number of queries changed: _____

3. Number of examined records: _____

4. Search time: _____

5. Comment:

2. Survey questionnaire 2.1. Email advertisement for survey

Dear____

The purpose of this message is to request information from you on a topic of concern to many in our field: access to information written in non-English languages. I am a doctoral student in library and information science at Rutgers University working on my dissertation. My doctoral committee includes Professors Dan O'Connor (Chair), Carol Kuhlthau, Tefko Saracevic, and, from OCLC, Dr. Lynn Silipigni Connaway. My dissertation includes two different methods, an experiment and a survey. This request deals with the questionnaire for the survey component of my study.

First, I am asking you for permission to post a brief message to your listserv: [name of listserv goes here]. That brief message is printed below. It includes the URL for a questionnaire available on the Internet through SurveyMonkey. Individuals who read my message can elect to participate or not participate in my research. Note in my sample message below that Rutgers has approved this study.

Thank you for considering my request to have a message posted on your listserv. I do hope that you will allow your subscribers the opportunity to participate in a study which addresses the importance of access to information on a global level.

Sincerely,

YooJin Ha

*** *** *** *** *** *** ***

Listserv Message Provided Below: *** *** *** *** *** ***

Dear Colleague:

The purpose of this message is to request information from you on a topic of concern to many in our field: access to information written in non-English languages. I am a doctoral student in library and information science at Rutgers University working on my dissertation. My doctoral committee includes Professors Dan O'Connor (Chair), Carol Kuhlthau, Tefko Saracevic, and, from OCLC (Online Computer Library Center), Dr. Lynn Silipigni Connaway. My dissertation includes two different methods, an experiment and a survey. This request deals with the questionnaire for the survey component of my study.

The purpose of my study is to explore non-English information seekers' information needs and study their information seeking behavior, especially users who use information

in non-Roman alphabet languages, such as Chinese, Japanese or Korean. It also includes exploring if new features are needed to improve cross-language access to bibliographic records in Databases and Online Public Access Catalogs for non-English information seekers. Additionally, it will attempt to determine the appropriateness of the bibliographic record.

In the questionnaire, you will be asked some questions about your experiences using information retrieval systems. You will be also asked to evaluate bibliographic records which are taken from online databases and online library catalogs. There are no correct answers to this questionnaire. The answers will merely reflect your assessments of the bibliographic records. Completing the questionnaire should take approximately 15 to 25 minutes.

The results of this survey will be reported anonymously. The information will be kept confidential by limiting individual's access to the research data and keeping it in a secure location. The researcher, her advisor, and the Institutional Review Board at Rutgers University are the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at a professional conference, only group results will be stated, unless you have agreed otherwise.

If you would like any information regarding this study, please feel free to contact YooJin Ha or Professor Dan O'Connor by e-mail or phone.

Sincerely,

YooJin Ha Ph.D. Candidate School of Communication Information and Library Studies Rutgers University 4, Huntington Street, New Brunswick, NJ, 08901, U.S.A. <u>yha@scils.rutgers.edu</u> Phone/email for Professor Dan O'Connor: 732-932-7500 x8219 <u>oconnor@scils.rutgers.edu</u>

2.2. Survey questionnaire

A Survey on Access to non-English Information

1. Consent Form

Consent Form for survey on "Access and Use of Non-English Information: Exploring User Issues with Multi-Language Database Systems"

You are invited to participate in a research study that is being conducted by YooJin Ha, the Principal Investigator of this study, who is a doctoral candidate in the department of School of Communication, Information, and Library Studies (SCILS) at Rutgers University. SCILS at Rutgers University supports the practice of protection for human subjects participating in research. The following information is provided for you to decide whether you wish to participate in the present study.

The purpose of this research is to explore non-English information seekers' information needs and study their information seeking behavior. It also includes exploring if new features are needed to improve cross-language access to bibliographic records in Databases and Online Public Access Catalogs for non-English information seekers. Additionally, it will attempt to determine the appropriateness of the bibliographic record.

You will be asked some questions about your experiences using information retrieval systems. You will be also asked to evaluate bibliographic records which are taken from online databases and online library catalogs. There are no correct answers to this questionnaire. The answers will merely reflect your opinions of the bibliographic records. Completing the questionnaire should take approximately 15 to 25 minutes.

The results of this survey will be reported anonymously. Anonymous means that the principal investigator, YooJin Ha will record no information about you that could identify you. This means that the Principal investigator will not record your name, address, phone number, date of birth, etc. The information will be kept confidential by limiting individual's access to the research data and keeping it in a secure location. The researcher, her advisor, Professor Dan O'Connor, and the Institutional Review Board at Rutgers University are the only parties that will be allowed to see the data, except as may be required by law. If a report of this study is published, or the results are presented at a professional conference, only group results will be stated, unless you have agreed otherwise.

Although no discomfort is anticipated while participating in this study, you are free to stop your participation by not completing the questionnaire if any of the questions make you feel uncomfortable at any time.

If you have any questions about your rights as a research subject, you may contact the Sponsored Programs Administrator at Rutgers University at: Rutgers University Institutional Review Board for the Protection of Human Subjects Office of Research and Sponsored Programs 3 Rutgers Plaza, New Brunswick, NJ 08901-8559 Tel: 732-932-0150 ext. 2104 Email: humansubjects@orsp.rutgers.edu

If you would like additional information regarding this study, including results, before or after its completion, please feel free to contact YooJin Ha by e-mail or Professor O'Connor by phone or email.

Sincerely,

YooJin Ha Doctoral candidate School of Communication Information and Library Studies Rutgers University 4, Huntington Street, New Brunswick, NJ, 08901, U.S.A. yha@scils.rutgers.edu

Phone/email for Professor Dan O'Connor 732-932-7500 x8219 oconnor@scils.rutgers.edu

A Survey on Access to non-English	Inform	ation					
1. Do you agree to participate t	his surv	iev?					
	ins surv	cy:					
() TES		O N	0				
	_	_	_	_	_	_	_
2. Online Searching Experie	ence						
Please note that you can go back and upda	te existin	g respon	ses until	the surve	y is finish	ed. After	finishing
the total survey, you will not be able to re-	enter the	e survey.					
1 How long ago did you last use	a lihra	urv to s	earch f	or infor	mation	in the	library
catalog or in any bibliographic of	latabas	es?	curchi	davs ac	io from	today	(if a
year or more, please indicate).							(
	c	<i>с</i>					-
2. Please provide your current u	ise of t	he follo	owing o	nline so	ources u	ising ti	ne /
point scale below.							
	Not at all			Moderate use			Most use
Library catalog (OPAC) (for example, Rutgers University Libraries System)	C	0	C	С	C	С	C
Worldwide union catalog systems (for example, WorldCat or RLIN etc.)	0	0	0	С	C	С	O
Online database for journals (for example, Web of Science, PsycINFO, ERIC, Medline etc.)	0	0	0	C	C	C	O
Web search engines (for example, Google, Yahoo etc.)	O	\odot	C	О	C	\odot	0
Web translation services (for example, Altavista Babel, Systran, WorldLingo etc.)	C	C	0	С	С	С	C
3. Please provide the names (or	r appro	ximate	names	s) of otl	ner syst	ems yo	ou may
use when searching for online i	nforma	tion (i.	e., spec	ial data	abases,	image	
databases, global positioning da	atabase	es, nun	nerical	latabas	ses, etc.	.)	
C None							
Names of other systems		_					
4. Blooco provido the name of the	ha liku-	m	co onli	no inf.	mation	Veller	o the
most often. Please provide the name of the	ibrarv I	ry who locatio	n such	as New	Brunsv	vick. N	J.
O None						, 11	-
Name and location of library							
							Page 2



A Survey on Access to non-English Information

3. Non-English information searching

This section asks about your non-English information searching experiences. Non-English information searching includes searches conducted in any languages other than English. Note that non-English information searching includes your native language if that language is not English.

Please note that you can go back and update existing responses until the survey is finished. After finishing the total survey, you will not be able to re-enter the survey.

1. Could you please indicate under what circumstances you might need to access or use information written in NON-ENGLISH languages?



2. Could you please indicate when was the last time you needed to access information in NON-ENGLISH languages? _____ days ago from today (if never or more than a year ago, please state so).

3. Could you please indicate when was the last time you needed to access information in languages that you cannot read? What was the language? Why did you need that? (if never or more than a year ago, please state so)

days ago from today	
Language you searched:	
Parana	
Reasons:	

4. Please provide your use of and evaluation of the following sources when you need NON-ENGLISH information including languages you do not know well (i.e. you would want to get information on Japanese animation written in Japanese although you don't know Japanese). Please indicate your USE of each system AND provide your overall evaluation of that source searching for NON-ENGLISH information using a letter grade.

A Survey on Access to non-English Information Overall System use evaluation · • • • • Library catalog (OPAC) (for example, Rutgers University Libraries System) ---Worldwide union catalog systems (for example, WorldCat or RLIN) - -Online database for journals (for example, EBSCO, Medline etc.) - -**- -**Web search engines (for example, Google, Yahoo etc.) - -Other (please specify the system name and evaluation) . -5. For EACH category below, please indicate the specific system name which you use when you search for NON-ENGLISH information. Library or library catalog (OPAC) Worldwide union catalog systems Web search engines Online database for journals Other systems 6. When you search non-English information, what information format are you usually seeking? Moderate Not at all Most use use Academic journal articles \odot O $^{\circ}$ \odot 0 \odot \odot News/magazine articles 0 C Books \odot 0 \odot \bigcirc 0 \bigcirc 0 Internet resources Media (movie, music etc) Other (please specify the format and your use level) . . 7. Please evaluate the following systems for searching materials in different format when you need information written in NON-ENGLISH including languages you do not know well. You can mark more than one format for each source. Academic journal News/magazine Media (movie, Internet Books articles articles resources music etc) Library catalog (OPAC) (for example, Rutgers University Libraries System)





A Survey on Access to non-English Information

4. Bibliographic records evaluation

Please note that you can go back and update existing responses until the survey is finished. After finishing the total survey, you will not be able to re-enter the survey.

** Using the numbered questions below, please evaluate the following bibliographic record.

More Like Thi	· Search far versions with same title and suther LAdvanced entires
Tiel	· Dearch for versions with same the and author providenced options
1100	e Paemjango suť yu oe :
	2002-yon che 26-hoe Yi Sang Munhaksang susang chakp`umjip /
Author(s	;: <u>Kwon, Chi-γe,; 1960-</u> ; Paemjango sut`yu.
Publication	n: Soul T`ukpyolsi∴ Munhak Sasangsa, Edition: Ch`op`an
Yea	2002
Description	; 380 p. # ill., ports. # 23 cm.
Language	: Korean (Show non-Roman characters)
Serie	;: Yi Sang Munhaksang chakp`umijp ;; che 26-hoe (2002); Variation: Yi Sang Munhaksang susan
Content	:: Paemjango sut`yu / Kwon Chi-ye Nae kasum e tchikhin sae ui palchaguk / Kwon chi-ye Ch`c Kim In-suk lin soguk / Yun Yong-su Chugun saram ui uibok / Chong Yong-mun Mari ui chip Un-yong Yoin / Han Ch`ang-hun Yuryong ui chip / Ch`oe In-ho.
Standard No	: ISBN: 8970124039; 9788970124032
	SUBJECT(S)
Descripto	Short stories, Korean 21st century, Korean fiction 21st century.
Note(s	: Includes bibliographical references.
Class Descriptor	:: LC: <u>PL981.7</u> , Dewey: <u>895.7/35</u>
Responsibility	r: [chiuni Kwon Chi-ye oe].
Material Type	v: Fiction (fic)
Document Type	:: Book
Entry	:: 20020403
Update	: 20070314
ð snarslan ðir	• OCI C 40517000

nun Darits VOU O	a not understand
,,	o not understand.
s that you erstand:	
s you do not	
erstand:	
What kinds of fu	inctions or services might be included to increase your
derstanding of t	the above record?
cord. If the abovely is it that you tential selection	ve record were retrieved from a search you conducted, how would examine it more carefully in order to consider for your ?
	not likely somewhat very like
would examine this reco	
Using the numl bliographic recor	pered questions below, please evaluate the following d.
Using the numb liographic recor	pered questions below, please evaluate the following rd.
Using the numl liographic recor	pered questions below, please evaluate the following rd. 赤地之戀 /
Using the numb Diographic record Customet and Title:	pered questions below, please evaluate the following rd. 赤地之感 / Chi di zhi lian /
Using the numb bliographic record custome units in record Title: Author(s):	bered questions below, please evaluate the following rd. 赤地之戀 / Chi di zhi lian / 張愛玲
Using the numb pliographic record contents and in record Title: Author(s):	bered questions below, please evaluate the following rd. 赤地之戀 / Chi di zhi lian / 張愛玲. <u>Zhang, Ailing.</u>
Using the numb Diographic record Title: Author(s): Publication:	pered questions below, please evaluate the following rd. 赤地之戀 / Chi di zhi lian / 張愛玲, Zhang, Ailing. 台北市;皇冠文學出版有限公司, 1991, Edition: 典藏版初版.
Using the numb pliographic record Title: Author(s): Publication:	bered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲. Zhang, Alling. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Talbei Shi: Huang guan wen xue chu tan you xian gong si.
Using the numb Diographic record Control of the control Title: Author(s): Publication:	bered questions below, please evaluate the following d. 赤地之感 / Chi di zhi lian / 張愛玲, Zhang, Alling. 台北市 : 皇冠文寧出版有限公司, 1991, Edition: 典藏版初版, Taibel Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban.
Using the numb pliographic record Title: Author(s): Publication: Year:	bered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲, Zhang, Ailing. 台北市:皇冠文學出版有限公司, 1991, Edition: 典藏版初版, Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991
Using the numb pliographic record Title: Author(s): Publication: Year: Description:	pered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲. Zhang, Alling. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibel Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p. ; 21 cm.
Using the numb pliographic record Title: Author(s): Publication: Year: Description: Language:	pered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 强愛玲. Zhang, Ailing. 台北市;皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibel Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p. ; 21 cm. Chinese (<u>Hide non-Roman characters</u>)
Using the numb pliographic record Title: Author(s): Publication: Year: Description: Language: Series:	bered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲, Zhang, Ailing. 台北市 : 皇冠文學出版有限公司, 1991, Edition: 典藏版初版. Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p.; 21 cm: Chinese (Hide non-Roman characters) 皇冠叢書; 第1891種
Using the numb pliographic record Title: Author(s): Publication: Year: Description: Language: Series:	bered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲. Zhang, Ailing. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p.; 21 cm. Chinese (Hitle non-Roman characteus) 皇冠叢書; 第1891種 張愛玲全集; 2
Using the numb pliographic record Title: Author(s) Publication: Year: Description: Language: Series:	pered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 强愛玲. Zhang, Ailing. 台北市:皇冠文學出版有限公司, 1991, Edition: 典藏版初版. Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 264 p.; 21 cm. Chinese (Hite non-Roman characters) 皇冠叢書;第1891種 張愛玲全集; 2 張愛玲. Works, 1991; 2.
Using the numb bliographic record Title: Author(s) Publication: Year: Description: Language: Series:	bered questions below, please evaluate the following d. 赤地之戀 / Chi di zhi lian / 張愛玲. Zhang, Ailing. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p.; 21 cm. Chinese (Hide non-Roman characters) 皇冠叢書;第1891種 張愛玲全集; 2 張愛玲. Works. 1991; 2. Huang guan cong shu ;; di 1891 zhong; Zhang Ailing quan ji ;; 2; Variation: Zhang, Ailing; J ISBN: 9672915402 :: 972955402
Using the numl pliographic record Title: Author(s) Publication: Year: Description: Language: Series: Standard No:	ered questions below, please evaluate the following d. 亦地之戀 / Chi di zhi lian / 張愛玲. Zhang, Alling. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibei Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991 254 p. ; 21 cm: Chinese (<u>Hite non-Roman characters</u>) 皇冠叢書;第1891種 張愛玲全集; 2 張愛玲. Works. 1991; 2. Huang guan cong shu ;; di 1891 zhong, Zhang Ailing quan ji ;; 2, Variation: Zhang, Ailing,; ISDN: 9573305402 ;; 9789573305408
Using the numl bliographic record Title: Author(s) Publication: Year: Description: Language: Series: Standard No:	pered questions below, please evaluate the following d. 赤地之感 / Chi di zhi lian / 張愛玲. Zhang, Alling. 台北市:皇冠文學出版有限公司, 1991. Edition: 典藏版初版. Taibel Shi : Huang guan wen xue chu ban you xian gong si, Edition: Dian cang ban chu ban. 1991. 254 p.; 21 cm. Chinese (Hitse non-Roman characters) 皇冠叢書; 第1891種 張愛玲全集; 2 張愛玲 Works. 1991; 2. Huang guan cong shu ;; di 1891 zhong; Zhang Alling quan ji ;; 2; Variation: Zhang, Alling.; ISBN: 9673305402 ;; 9789673305408 SUBJECT(5)

	o non-Englisł	n Informa	ation					
Class Descriptors:	LC: PL2837.E35							
Responsibility:	[作者張愛玲]. [zuo zhe Zhang Ai	ling].						
Material Type:	Fiction (fic)							
Document Type:	Book							
Entry:	19910911							
Update:	20070314							
Accession No:	OCLC: 24360139							
5. Please indicate erms of its conte	the degree to nt.	o which	you unde	erstar	nd the al	oove re	cord i	N completely
	L	understand		u	nderstand			understand
nderstand the record		O	O	0	O	C	O	0
. What kinds of fu	inctions or se	ervices n	night be i	ncluc	led to in	crease	your	
inderstanding of	the above re	cord?						
3. Please assume record. If the abo ikely is it that you potential selection	you need info ve record we would exam ?	ormation re retrie ine it mo	about th eved fron pre carefu	ie top 1 a se Illy in	bic inclue arch yo a order t	led in tl u condu o consic	he abo icted, ler foi	ove how r your
3. Please assume record. If the abo ikely is it that you potential selection	you need info ve record we would exam ?	ormation re retrie ine it mo	about th eved fron ore carefu	ie top 1 a se Illy ir	oic inclue carch you n order te somewhat likely	led in tl u condu o consic	he abo icted, ler foi	ove how r your very likely
8. Please assume record. If the abo likely is it that you potential selection 'ou would examine this reco	you need info ve record we would exam ? *	ormation re retrie ine it mo	about the eved from ore carefu	e top 1 a se 111y ir	pic inclue earch you order to somewhat likely	ded in th u condu o consid	he abo icted, ler for	ove how r your very likely

Mue Like This. Title:	Auvanceu opuons
Title:	
	日米関係におけるエスニシティーの要素 = Ethnicity in
	U.SJapan relationship.
	Nichi-Bei <mark>kankei</mark> ni okeru esunishitī no yōso =
	Ethnicity in the history of the U.SJapan relationship.
Corp Author(s)	総合研究開発機構 (Japan)
	Sõgõ Kenkyū Kaihatsu Kikõ (Japan)
Publication:	東京:総合研究開発機構:総発売元全国官報販売協同組合,平成7[1995
	Tōkyō : Sōgō Kenkyū Kalhatsu Kikō : Sōhatsubaimoto Zenkoku Kanpō Hanbai Kyōdō Kumia
Year.	1995
Description:	ix, 177 p. ; 26 cm.
Language:	Japanese (Hide non-Roman characters), Text in Japanese, summary and table of contents also in
Series:	NIRA 研究報告書 = NIRA research report ; no. 940052
	NIRA 研究報告書; no. 940052.
and a state of the state of the	NIRA kenkyu nokokusno -, NIRA research report ,, no. 940052, Variation. NIRA kenkyu no
Standard No.	ISBN: 4795514208; 9784795514201 LCCN: 96-459429
	SUBJECT(S)
Descriptor	Ethnicity – United States
Geographic	Linited States – Relations – Janan
ocographic.	Japan – Relations – United States
Note(s):	Cover title./ Includes bibliographical references.
Class Descriptors:	LC: GN495.6
Other Titles:	Ethnicity in the history of the U.SJapan relationship
Material Type:	Government publication (gpb); National government publication (ngp)
Document Type:	Book
ne	

A Survey on Access to non-English Information parts you do not understand:				
11. What kinds of functions or services might be included to increase your understanding of the above record?				
12. Plea record. likely is potentia	ase assume you need inform If the above record were re it that you would examine it al selection?	ation about the strieved from a more carefully	topic included in search you cond in order to consi	the above ucted, how der for your
You would ex	xamine this record	<u> </u>	C C	0
** Using bibliogr	g the numbered questions b aphic record.	elow, please ev	aluate the follow	ing
More Like This:	Search for versions with same title and author A	dvanced options		
Title:	: Seals of Chinese painters and collecto	ors of the Ming and C	h`ing periods,	
Author(s):	: <u>Wang, Chi-ch`ien.</u> <u>Contag, Victoria,</u> ; joint comp.			
Publication:	: [Hongkong] Hong Kong University Press, Edition: Rev. ed. with supplement.			
Year:	: 1966			
Description:	: Ixviii, 726 p. illus. 27 cm.			
Language:	: Chinese (Show non-Roman characters)			
Standard No:	: LCCN: 66-84760			
	SUBJECT(S)			
	Artists' marks. Collectors' marks. Painters Dictionaries German. Painters Dictionaries English. Art Collectors and collecting. Sceaux Chine. Artistes Signatures. Collections Marques. Peintres Dictionnaires allemands. Peintres Dictionnaires anglais. Art Collectionneurs et collections. Peintres Chine.			
Note(s):	: First ed. published in 1940 under title: Ming Qing Prefatory matter in English and Chinese, text in C yin jian./ "Seals of Sung and Yüan painters"; p. 5 American public and private collections"; p. [631]-	hua jia yin jian and Wang's <mark>hinese</mark> , with translations in 13-528./ "American supplen 726./ "Source material": p.	name in characters appeared German or English; added t.p. nent: further <mark>seals</mark> found on <mark>Ch</mark> [xxii]-xxiii. "Literature": p. xxvii	on the title page./ : Ming Qing hua jia <mark>inese</mark> paintings in -xxviii.
ass Descriptors:	: LC: <u>N45;</u> Dewey: <u>929.8</u>			
Other Titles:	: Ming Qing hua jia yin jian.			
Responsibility:	: reproduced in facsimile size and deciphered, by \searrow	/ictoria Contag and Wang C	hi-ch`ien. Introd. by James Ca	hill.
More Records:	Show record information			
Document Type:	: Book			
Entry:	: 19740430			
Update:	: 20070407			
Accession No:	: OCLC: 872730			
Database:	: WorldCat			
13. Please erms of it	e indicate the degree to whi ts content.	ch you underst	and the above r	ecord in
nderstand the r	understand		understand	understand
iverstand the n				

14. Please indicate what parts of the above record that you understand, and which parts you do not understand.

A Survey on Access to non-En	alish Information		
parts that you	5		
understand:			
understand:			
15. What kinds of functions	or services might	t be included to increa	se your
understanding of the above	e record?		
16. Please assume you nee	d information abo	ut the topic included i	in the above
record. If the above record	were retrieved f	rom a search you con	ducted, how
likely is it that you would ex	xamine it more ca	refully in order to con	sider for your
potential selection?			
	not likely	somewhat likely	very likely
You would examine this record	0 0	<u> </u>	0 0
5. Background question	IS		
Please note that you can go back and	l update existing respor	nses until the survey is finish	ned. After finishing
the total survey, you will not be able	to re-enter the survey	. ,	5
1 What is your native lang	uade2		
	uugei		
2. Where were you born?			
3. If you were not born in t	he US, how long h	ave you lived in the U	IS?
years			
4. Except for your mother l	anguage, please i	ndicate your knowled	lge of other
languages if you have. Usin	ig a 100% scale, i	ndicate your reading	
comprehension level, speak	king level, and wri	ting level for that land	mage A score
	<i>a i i i i i i i i i i</i>	ang let er ter anat tang	gauger A score

A Survey on Access	s to non-English Information
* Other Language (fill in)	
Reading: %	
Speaking: %	
Writing: %	
* Other Language (fill in)	
Reading: %	
Speaking: %	
Writing: %	
* Other Language (fill in)	
Reading: %	
Speaking: %	
Writing: %	
5 Dlease specify	additional languages you know other than those listed in the
two questions a	bove.
Languages:	
6 Blasca list you	r academic degrees. Blease specify all the college/university
degrees you bay	in academic degrees. Please specify an the conege/university
uegrees you nu	
* Degree	
Major	
Year	
* Degree	
Major	
Year	
* Degree	
Major	
Year	
* Degree	
Major	
Year	
7. Please provid	e the title or approximate title of a few well-known scholarly
	Page 15

A Survey on Access to non-English Information journals in your area? ▲ \mathbf{v} 8. Among your recommended journals, is there any journal written in a language other than English? If yes, please indicate its title or approximate title. If no, please provide your assessment as to why there is no scholarly recognized journal written in English in this area? Yes - indicate the journal title or approximate title No - provide your assessment the reason why major journals in your area are primarily in English 9. Please provide the title of your current position. Be as specific as possible (for example, professor, Asian studies, Rutgers University; librarian, Asian language cataloger, Rutgers University; doctoral student, education major, Rutgers University). -Ψ. 10. How many years of experience have you had in your current position? _ years. 11. What is the total number of years of work experience you have had before you got your current position? _____ years. 12. What is your gender? Page 16

A Survey on Access to non-English Information

6 Female

O Male

6. Thank you!

Thank you very much for completing this survey! It is hoped that your responses will contribute to our knowledge of the access needed for bibliographic records, especially when accessing non-English information.

If you would like to know additional information regarding this study, please feel free to contact YooJin Ha by e-mail or Professor O'Connor by phone or email.

Sincerely,

YooJin Ha Doctoral candidate School of Communication Information and Library Studies Rutgers University 4, Huntington Street, New Brunswick, NJ, 08901, U.S.A. yha@scils.rutgers.edu

Phone/email for Professor Dan O'Connor 732-932-7500 x8219 oconnor@scils.rutgers.edu

Bibliography

- Abbas, J. (2005). Out of the mouths of middle school children: I. Developing userdefined controlled vocabularies for subject access in a digital library. *Journal of the American Society for Information Science and Technology*, *56*(14), 1512-1524.
- Aliprand, J.M. (1999). Cataloging in the Universal character set environment: Looking at the limits. In 65th IFLA Council and General Conference, Bangkok, Thailand, August 20-Augst 28, 1999
- Allan, R. B. (1990). User models: theory, method, and practice. *International Journal Man-Machine Studies*, *32* 511-543.
- Anderson, J. D. (1974). Foreign language barriers in information transfer. *Journal of Education for Librarianship*, 14(3), 171-184.
- Arnold, L., Sadler, L., Balkan, S. M., & R. L. Humphreys (1997). *Machine Translation: an Introductory Guide*. London: Blackwells.
- Arsenault, C. (2005). Retrieval Issues for Chinese-Language Library Resources: A North American Perspective. *The International Information & Library Review*, 37(2), 107–116.
- Barry, C. L. & Schamber, L. (1998). Users' Criteria for Relevance Evaluation: A Crosssituational Comparison. *Information Processing and Management.* 34(2/3), 219-236.
- Barry, C. L. (1994). User-defined relevance criteria: an exploratory study. *Journal of the American Society for Information Science and Technology*, 45(3), 149-159.
- Bates, M. J. (1991). The berry-picking search: User interface design. In M. Dillon (Ed.), Interfaces for information retrieval and online systems: The state of art (pp. 55-61). New York: Greenwood Press.
- Bates, M. J. (1990). Where Should the Person Stop and the Information Search Interface Start?. *Information Processing & Management 26*, 575-591.
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, *13* (5), 407-424.
- Bates. M.J. (1979). Information search tactics. *Journals of the American Society for Information Science*, 30(4), 205-214.
- Bates, M. J, Wilde, D., & Siegfried, S. (1993, January). An Analysis of Search Terminology Used by Humanities Scholars: The Getty Online Searching Project Report No. 1. *Library Quarterly* 63, 1-39.
- Ballesteros, L., & Croft, W. (1998). Resolving ambiguity for cross-language retrieval. In Processing of SIGIR Conference, 64-71.
- Beitzel, S., Jensen, C., Chowdhury, A., Grossman, D., & Frieder, O. (2004). In Hourly Analysis of a Very Large Topically Categorized Web Query Log (pp. 321-328).
 Paper presented at *the ACM SIGIR Conference* on Re-search and Development in Information Retrieval, Sheffield, UK. ACM Press.
- Belkin, N. J. (1996). Intelligent information retrieval: Whose intelligence? *Proceedings* of the Fifth International Symposium for Iinformation Science (pp. 25–31). Konstanz: Universitä⁻⁻tsverlag Konstanz.
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, *5*, 133-143.

- Belkin, N. J., C. Cool, A. Stein & U. Thiel. (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 system: A review of research from document retrieval to knowledge-based system. London: The British Library.
- Belkin, N.J., Oddy, R.N. & Brooks, H. (1982a). ASK for Information retrieval Part I. *Journal of Documentation*, 38 (2), 61-72.
- Belkin, N.J., Oddy, R.N. & Brooks, H. (1982b). ASK for Information retrieval Part II. *Journal of Documentation*, 38 (3), 145-164.
- Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science 5*, 133-143.
- Bernstein, J. H. (2006). From the ubiquitous to the Nonexistent: a Demographic Study of OCLC WorldCat. *Library Resources & Technical Services*, *50*(2), 79-90.
- Berger, M. (1999). An annotated guide to Internet resources on immigration and immigrants. *SLA Social Science Division Bulletin*, 40(3), 5.
- Bilal, D. (2000). Children's use of Yahooligans! web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science and Technology*, *51*, 646-665.
- Borgman, C. L (1997, June). Multi-Media, Multi-Cultural, and Multi-Lingual Digital Libraries Or How Do We Exchange Data in 400 Languages? *D-Lib Magazine, June*.
- Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons learned from information retrieval studies. *Journal of the American Society Information Science*, 37(6). 387-400.
- Borgman, C. L. (1984). Psychological research in human-computer interaction. In *Annual Review of Information Science and Technology*, *19*, 33-64.
- Borgman, C. L. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*, 47, 493-503.
- Borgman, C. L., Hirsh, S., & Hiller, J. (1996). Rethinking online monitoring methods for information retrieval systems: from search product to search process. *Journal of the American Society Information Science*, 47(7), 568-583.
- Borgman, C. L. & Siegfried, S. L. (1992). Getty's Synoname and Its Cousins: A Survey of Applications of Personal Name-Matching Algorithms. *Journal of the American Society Information Science*, 43(7), 459-476.
- Bossmeyer, C.; Massil, S.W. (Eds.). (1987). Automated systems for access to multilingual and multiscript library materials : problems and solutions : papers from the preconference held at Nihon Daigaku Kaikan Tokyo, Japan, August 21-22, 1986. International Federation of Library Associations and Institutions, Section on Library Services to Multicultural Populations and Section on Information Technology. Munich and New York: K.G. Saur.
- Case, D. (2002). Looking for information: a survey of research on information seeking, needs, and behavior. San Diego, CA: Academic Press.
- Chan, J. (1995). Evaluation of formats used to display bibliographic records in OPACs in Canadian Academic and public libraries. Master of Information Science research project report (University of Toronto: Faculty of information studies).

- Chen, H. & Dhar, V. (1991). Cognitive process as a basis for intelligent retrieval systems design. *Information Processing and Management*, 27(5), 405–432
- Cheng, Y. (2004). *Thoughts, feelings, and actions: quantitative comparisons of interactions and relationships among three factors in college students' information seeking.* (Doctoral Dissertation, Indiana University, 2004).
- Cherry, J., Muter, P. & Szigeti, S. (2006). Bibliographic displays in web catalogs: does conformity to design guidelines correlate with user performance? *Information technology and libraries*, *25*(*3*), 154-162.
- Cherry, J. & Cox, J. (1996). World Wide Wed Displays of bibliographic records: an evaluation. *Proceedings of the 24th conference of the Canadian Association for Information Science*, Toronto, Ontario: Canadian Association for Information Science, 101-114.
- Chowdhury, G. (2003). Natural Language Processing. Annual Review of Information Science and Technology (ARIST), 37, 51-89.
- Chung, W., Zhang, Y., Huang, Z., Wang, G., Ong, T-H., & Chen, H. (2004). Internet searching and browsing in a multilingual world: An experiment on the Chinese Business Intelligence Portal (CBizPort). *Journal of the American Society for Information Science and Technology*, 55(9), 818-831.
- Cole, J. C. & Cole, C. (1972). The Ortega Hypothesis. *Science, New Series*, 178(4059), 368-375.
- Cooper, M. D. & Chen, H-M. (2001). Predicting the relevance of a library catalog search. *Journal of the American Society for Information Science and Technology*, 52(10), 813-827.
- Cosijn, E. & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing* and Management, 36, 533-550.
- Carpenter, M. P & Narin, F. (1981). The adequacy of the science citation index (SCI) as an indicator of international scientific activity. *Journal of the American Society for Information Science*, 32(6), 430-439.
- Davis, J. L. & Livny. (1994). Monitoring Japanese scientific and technical information using JICST databases. *Database Magazine*, *17*(*3*), 33-42.
- Davis, M., & Ogden, W. (1997). QUILT: Implementing a Large-Scale Cross-Language Text Retrieval System. In Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, July 1997, 92-98.
- De Mey, M., R. Pinxten, M. Porian, & F. Vandamme. (Eds.). (1977). *International Workshop on the Cognitive Viewpoint. 1977 March 24-36*, Ghent University, Belgium.
- De Mey, M. (1980). The Relevance of the Cognitive Paradigm for Information Science. In Ole Harbo & Kajberg (Eds.), *Theory and Application of Information Research Proceedings of the 2nd International Research Forum on Information Science1977, August 3-6, Copenhagen, Denmark, Royal School of Librarianship* (pp. 48-61). London, England: Mansell.
- De May, M. (1982). *The Cognitive Paradigm: An Integrated Understanding of Scientific Development*. Dordrecht, Netherland: D. Reidel Publishing co.
- Dervin, B. (1999). On studying information seeking methodologically: the implications of connecting metatheory to method. *Information Processing and management*,

35, 727-750.

- Dervin, B. (1992). From the mind's eye of the user: The Sense-Making qualitativequantitative methodology. In J. D. Glazier & R. R. Powell (Eds.), *Qualitative research in information management* (pp. 61-84). Englewood, CO: Libraries Unlimited.
- Dervin, B. (1983) An overview of sense-making: concepts, methods and results to date. Paper presented at the *International Communication Association Annual Meeting*, Dallas TX, May 1983.
- Dervin, B., & Nilan, M. (1986). Information needs and uses. In Annual Review of Information Science and Technology, 21, 3-33.
- Echabe, A.E. & Castro, J.L. (1998). Social memory: macropsychological aspects. In U. Flick (Eds.), *The Psychology of the Social* (pp. 91-106). UK: Cambridge University Press.
- Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, *63(4)*, 469–486.
- Ellis, D. (1989). A behavioural approach to information retrieval design. *Journal of Documentation*, 46, 318-338
- Ellis, D. Allen, D. & Wilson, T. (1999). Information Science and Information Systems: Conjunct Subjects Disjunct Disciplines. *Journal of the American Society for Information Science and Technology*, 50(12), 1095-1107.
- Fidel, R. (1985). Moves in online searching. Online Review, 9, 61-74.
- Fidel, R. (1991). Searchers' selection of search keys: I. The selection routine, II. Controlled vocabulary or free-text searching, and III. Searching styles. *Journal of the American Society for Information Science*, 42, 490-527.
- Fisher, K., Durrance, J. & Hinton, M. B. (2003). Information grounds and the use of need-based services by immigrants in Queens, New York: a context-based, outcome evaluation approach. *Journal of the American Society for Information Science and Technology*, 55(8), 754-766.
- Foster, A. (2004). A nonlininer model of information-seeking behavior. *Journal of the American Society for Information Science and Technology*, *55*(*3*), 228-237.
- Ganss, D. (1999). A nation of immigrants: Web sites on immigration. *The School Librarian's Workshop*, 20(3), 13-15.
- Gey, F., Kando, N. & Peters, C. (2005). Cross-Language Information Retrieval: the way ahead. *Information Processing and Management*, *41*, 415–443.
- Gonzalez, L. (1999). Public libraries reach out to new North Carolinians: Meeting the information needs of immigrants and refugees. *North Carolina Libraries*, *57*(*1*), 4-7.
- Grabois, H. (1999). The convergence of Scociocultural Theory and Cognitive Linguistics: Lexical Sementics and the L2 Acquisition of Love, Fear and Happiness. In G.B. Palmer & D.J. Occhi (Eds.), *Languages of Sentiment* (pp. 201-233). PA: John Benjamins.
- Grefenstette, G. (1998). The Problem of Cross-Language Information Retrieval. In Gregory Grefentte. (Ed.) (p.1-10). Cross-Language Information Retrieval. Norwell, MA : Kluwer.
- Ha, Y. (2006, June). Users' Information Seeking Behaviors With Cross Language Information Retrieval Systems. Paper presented at the Library Research

Roundtable session in the 2006 American Library Association, New Orleans, LA

- Ha, Y. & O'Connor, D. (2005, September). A Global Reach for the Information Disciplines. Paper presented at *i*-Conference, Penn State University.
- Ha, Y. (2005, January). *Beyond Language Barriers: Survey of current LIS Education and Online library system.* Paper presented at the Association for Library and Information Science Education Conference, Boston, MA.
- Hansen, P. & Karlgren, J. (2005). Effects of foreign language and task scenario on relevance assessment. *Journal of Documentation*, 61(5), 623-639.
- He, B., Patel, M., Zhang, Z., & Chang, K. C. (2007). Accessing the Deep Web; Attempting to Locate and Quantify Material on the Web that is Hidden from Typical Search Techniques. *Communications of the ACM*, 50(5), 95-101.
- Hjørland, B. (2002), Epistemology and the Socio-Cognitive Perspective in Information Science. *Journal of the American Society for Information Science and Technology*, 53(4), 257-270.
- Hjørland, B (1997). Information seeking and subject representation: An activitytheoretical approach to information science. Westport, CT; London: Greenwood Press.
- Hjørland, B. & Albrechtsen, H. (1995). Toward a new horizon in information science: domain analysis. *Journal of the American Society for Information Science*, 46(6), 400-425.
- Holzner, B. (1996). Improving access to global information resources {dissemination of foreign scholarly research}. In: Renewing the ARL agenda. Association of Res. Libs., 1996
- Hsieh-Yee, I. (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, 161-174.
- Hurych, J. (1986). After Bath: scientists, social scientists, and humanists in the context of online searching. *The Journal of Academic Librarianship*, *12*(*3*), 158-165.
- Hutchins, W. J., Pargeter, L.J., & Saunders, W.L. (1971). University research and the language barrier. *Journal of Librarianship*, 3(1), 1-25.
- Hyldegård, J. (2006). Collaborative information behaviour exploring Kulthau's Information Search Process model in a group-based educational setting. *Information Processing and Management*, 42, 276-298.
- Ingwersen, P., & Järvelin, K. (2005). *The turn: integration of information seeking and retrieval in context*. Dordrecht : Springer.
- Ingwersen, P. (1999). Cognitive Information Retrieval. Annual Review of Information Science and Technology (ARIST), 34, 3-51.
- Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 1996, 3-50.
- Ingwersen, P. (1992). Information Retrieval Interaction. London: Taylor Graham, 1992
- Ingwersen, P. (1982). Search procedures in the library analyzed from the cognitive point of view. *Journal of Documentation*, *38*, 165-191.
- Iivonen, M., & White, M. D. (2001). The choice of initial Web search strategies: A comparison between Finnish and American searchers. *Journal of Documentation*, 57, 465–491.
- Jacob, E. & Shaw, D. (1998). Sociocognitive perspectives on representation. Annual

Review of information Science and Technology (ARIST), 33, 131-185.

- Jiao, Q.G., & Onwuegbuzie A.J. (1997). Antecedents of library anxiety. *Library Quarterly*, 67, 372-389.
- Jansen, B., & Spink, A. (2006). How we are Searching the World Wide Web? A Comparison of Nine Searech Engine Transaction Logs. *Information Processing* and Management, 42(1), 248-263.
- Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users and real needs: a study and analysis of users' queries on the web. *Information Processing and Management*, *36*(2), 207–227.
- Jansen, B.J., & Pooch, U. (2001). A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science & Technology*, 52, 235-246.
- Jeong, W. (2004). Unbreakable ethic bonds: Information –seeking behavior of Korean graduate students in the United States. *Library Information Science Research*, *26*, 384-400.
- Jeong, W. (1998). A Pilot Study of OCLC CJK Plus as OPAC. *Library & Information Science Research*, 20, 3, 271-292.
- Jones, P.A. (1999). *Libraries, immigrants, and the American experience*. Westport, CT: Greenwood Press.
- Kellsey, C. (2003, June). Crisis in foreign language expertise in research libraries. *C&RL News*, 391-392, 397.
- Kelly, D. & Fu, X. (2007). Eliciting better information need descriptions from users of information search systems. *Information Processing and Management*, 43, 30-46.
- Kelly, G. A. (1963). *A theory of personality: the psychology of personal constructs*. New York: W.W. Norton.
- Kelly, J. (1992). Online in the tower of babble dropping the language barrier from databases access. P. 167-182. Book chapter.
- Kim, J.-A. (2006). Toward an understanding of web-based subscription database acceptance. *Journal of the American Society for Information Science*, *57*(*13*), 1715-1728.
- Kim, K.-S. (2002). Information seeking on the Web: Effects of user and task variables. *Library and Information Science Research*, *23*, 233–255.
- Kim, S. (2006). Romanization in Cataloging of Korean Materials. *Cataloging & Classification Quarterly*, 43(2), 53 76
- Kim, Y.-W. (2004). Typology of user of user uncertainty in Web-based information seeking: insight into the information seeking context of scholarly researchers in the field of science. (Doctoral Dissertation, Rutgers The State University of New Jersey, 2004).
- Krathwohl, D. (1998). Methods of educational & social science research: An integrated approach. New York, NY: Longman.
- Krikelas, J. (1980). Searching the Library Catalog--A Study of Users' Access. *Library Research*, 2(3), 215-230.
- Kuhlthau, C. C. (2005). Kuhlthau's Information Search Process. In Karen E. Fisher, Sandra Erdelez, and Lynne McKechnie (Eds.). *Theories of Information Behavior*. Medford, NJ: Information Today Inc.
- Kuhlthau, C. C. (2004). Seeking meaning: A process approach to library and information

services. Westport, CT: Libraries Unlimited.

- Kuhlthau, C. C. (1999a). The role of experience in the Information search Process of an early career information worker: perceptions of uncertainty, complexity, construction, and sources. *Journal of the American Society for Information Science*, 50(5), 399-412.
- Kuhlthau, C. C. (1999b). Accommodating the user's information search process: challenges for information retrieval system designers. *Bulletin of the American Society for Information Science, February/March*, 12-16.
- Kuhlthau, C.C. (1993a). Seeking meaning: A process approach to library and *information services*. Norwood, NJ: Alex publishing corporation.
- Kuhlthau, C.C. (1993b). A principle of uncertainty for information seeking. *Journal of documentation*, 49(4), 339-355.
- 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.
- Kuhlthau, C. C. (1988a). Longitudinal case studies of the information search process of users in libraries. *Library & Information Science Research*, *10*(*3*), 257-304.
- Kuhlthau, C. C. (1988b). Perceptions of the information search process in libraries: A study of changes fro high school through college. *Information Processing and Management*, 24(4), 419-428.
- Kuhlthau, C. C. (1988c). Developing a model of the library search process: Cognitive and affective aspects. *RQ*, 28, 232-242.
- Kuhlthau, C. C. (1983). The Library Research Process: Case Studies and Interventions with high School Seniors in Advanced Placement English Classes Using Kelly's Theory of Constructs. (Doctoral Dissertation, Rutgers University, 1983).
- Kuhlthau, C., Spink, A., & Cool, C. (1992). Exploration into stages in the information search process in online information retrieval. *Proceedings of the ASIS annual meeting*, *29*, 67-71.
- Lagier, J. (2002). Measuring usage and usability of online databases at Hartnell College: an evaluation of selected electronic resources. (Doctoral Dissertation, Nova Southeastern University, 2002).
- Large, A, &Moukdad, H. (2001). Multilingual access to web resources: An overview. *Program*, 34, p. 43–58.
- Lau, E. P. & Goh, D. H. (2006). In search of query patterns: A case study of a university OPAC. *Information Processing and Management*, 42(5), 1316-1329.
- Lavrenko, V., Choquette, M., & Croft, W.B. (2002). Cross-lingual relevance models. In: M. Beaulieu, R. Baeza-Yates, S.H. Myaeng and K. Järvelin (Eds). Proceedings of the 25th annual international ACM-SIGIR conference on research and development in information retrieval. (pp. 175–182) ACM Press, New York.
- Leuski, A., Lin, C., Zhou, L., Germann, U, Och, F., and Hovy, E. (2003). Cross-Lingual C*ST*RD: English Access to Hindi Information. *ACM Transaction on Asian Language Information Processing*, 2(3), 245-269.
- Liao, Y., Mary, F. & Lu, J. (2007).Information-seeking behavior of international graduate students vs. American graduate students: A user study at Virginia Tech 2005. *College and research libraries*, 68(1), 5-25.
- Liebergot, H. L. (1994). Use of international networks and databases to access United

States scientific information from Japan: a case study of Japan's national laboratory for high energy physics research. (Doctoral dissertation, Drexel University, 1994).

Littlejohn, S. (1992). Theories of human communication. Belmont, CA: Wadsworth.

Lincoln, Y., & Guba, E. (1985). Naturalistic inquiry. New York: Sage.

Lopez-Ostenero, F., Peinado, V., Gonzalo, J., & Verfjo, F. (2008). Interactive question answering: is cross-language harder than monolingual searching?. *Information*

Processing and Management, in press, 44(1), 66-81.

- Luria, A.R. (1979).*The making of Mind: A Personal Account of Soviet Psychology*. Cambridge: Harvard UP
- Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in fulltext end-user-oriented search systems: The roles of domain and search expertise. *Library & Information Science Research*, 15, 35-69.
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge England: Cambridge University Press.
- Markey, K. (1983). Online catalog use: results of surveys and focus group interviews in several libraries: Final report to the Council on Library Resources v. 2. Dublin, OH : OCLC Online Computer Library Center, Office of Research.
- Miles, M. & Huberman, A. M. (1994). *An expanded source book: Qualitative Data Analysis.* Thousand Oak, Ca: Sage publications Ltd.
- Mizzaro, S. (1997). Relevance: The Whole History. *Journal of the American Society for Information Science and Technology*, 48(9), 810-832
- Nahl, D. (2004). Measuring the affective information environment of web searchers. *Proceedings of the 67th ASIS&T Annual meeting (41),* 191-197.
- Nahl, D. & Tenopir, C. (1996). Affective and cognitive searching behavior of novice endusers of a full-text database. *Journal of the American Society for Information Science*, 47, 276-286.
- Oard, D. W. (1997). Serving users in many languages: cross-language information retrieval for digital libraries. *D-Lib Magazine*. Retrieved January 15, 2008, from <u>http://www.dlib.org/dlib/december97/oard/12oard.html</u>
- Oard, D.W. & Diekema, A.R. (1998). Cross-Language Information Retrieval. *Annual Review of Information Science and Technology (ARIST)*, 33, 223-256.
- Oard, D. W. & Resnik, P. (1999). Support for interactive document selection in cross language information retrieval. *Information Processing and Management*, 35, 363-379.
- Oard, D. & Peters, C. (1999). Multilingual Information Discovery and Access: An Introduction. Joint ACM Digital Library/SIGIR Workshop on Multilingual Information Discovery and AccesS (MIDAS). Retrieved on January 20, 2008 from http://www.clis2.umd.edu/conferences/midas/papers/midasintro.htm
- Oard, D., Ruiz, M., Klavans, J. Peters, C., Frederking, R. & Sheridan, P. (1999, October). Multilingual Information Discovery and Access (MIDAS). *D-Lib Magazine*, 5, Retrieved on January 20, 2008 from http://www.dlib.org/dlib/october99/10oard.html
- O'Connor, Daniel O. (1978). The Interactive Influences of Person and Situation Characteristics on Expectations and their Cumulative Effects on Relevance Judgments and File Preference Choices by Undergraduates Searching

Alphabetical and Classed Subject Catalogs. (Doctoral Dissertation, Syracuse University, 1978.)

- Ogden, D., Cowie, J., Davis, M., Ludovik, E., Molina-Salado, H. & Shin, H. (1999). Getting information from Documents You Cannot Read: An Interactive Crosslanguage Tex Retrieval and Summarization System. Joint ACM Digital Library/SIGIR Workshop on Multilingual information Discovery and AccesS (MIDAS).
- Ogden, D., Cowie, J., Davis, M., Ludovik, E., Nirenburg, S., Molina-Salgado, H., & Sharples., N. (1999) Keizai: An Interactive Cross-Language Text Retrieval System. Paper presented at the Workshop on Machine Translation for Crosslanguage Information Retrieval, Machine Translation Summit VII, September 13-17, 1999, Singapore.
- Ogden, D., & Davis, M. (2000). Improving Cross-Language Text Retrieval with Human Interactions. *Hawaii International Conference on System Sciences, HICSS-33*, January 4-7, 2000.
- Onwuegbuzie, A. J., & Jiao, Q. G. (1998). The relationship between library anxiety and learning styles among graduate students: Implications for library instruction. *Library & Information Science Research*, 20(3), 235-249.
- Orengo, V. & Huyck, C. (2006). Relevance feedback and cross-language information retrieval. *Information Processing and Management*, 42, 1203-1217.
- Park, J. (2001). Information retrieval of Korean materials using the CJK bibliographic system: issues and problems. *KSAA CONFERENCE*, 2001, 9.25.
- Perrault, A. H. (2003). The Role of WorldCat in Resources Sharing. *Collection Management*, 28(1/2), 63-75
- Peters, C. (2006). New directions in multilingual information access. *ACM SIGIR Forum*, 40(2), 31 39
- Peters, C. (2001). Cross-Language Information Retrieval (CLIR). DELOS Summer School on DL Technology. Lecture Notes. Retrieved on February 22, 2008 from <u>http://www.iei.pi.cnr.it/DELOS/delos2/SummerSchool/ handouts1/ISDL1.ppt</u>
- Peters, C. (Ed.) (2000). Cross-Languag Information Retrieval and Evaluation: Workshop of the cross-language evaluation forum, CLEF 2000, Lisbon, Portugal, September, 2000
- Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech*, 42(11:2), 41-66.
- Petrelli, D. (2008). On the role of user-centred evaluation in the advancement of interactive information retrieval. *Information Processing and Management, 44*, 22–38.
- Petrelli, D., Levin, S., Beaulieu, M., & Sanderson, M. (2006). Which user interaction for cross-language information retrieval? Design issues and reflections. *Journal of the American Society for Information Science and Technology*, *57*(*5*), 709-72.
- Petrelli, D., Beaulieu, M., Sanderson, M., Demetriou, G., Herring, P., & Hansen, P. (2004). Observing users, designing clarity: A case study on the user-centered design of a cross-language information retrieval system. *Journal of the American Society for Information Science and Technology*, 55, 923–934.
- Pettigrew, K., Fidel, R., & Bruce, H. (2001). Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology*, 35, 43-78.

- Pirkola, A., Puolamäki, D., & Järvelin, K. (2003). Applying query structuring in crosslanguage retrieval, *Information Processing and Management*, *39*(*3*), 391-402.
- Qin, J. Zhou, Y, Chau, M, & Chen, H. (2006). Multilingual Web retrieval: An experiment in English-Chinese business intelligence *Journal of the American Society for Information Science and Technology*, 57(5), 671-683
- Rayward, W. B. (1975). The Universe of Information: The Work of Paul Otlet for Documentation and International Organization. FID Publication 520. Moscow: Published for the International Federation for Documentation by the All-Union Institute for Scientific and Technical Information (VINITI), USSR Academy of Sciences.
- Rieh, S. Y. (2004). On the Web at home: Information seeking and Web searching in the home environment. *Journal of the American Society for Information Science and Technology*, 55, 743–753.
- Rieh, H. Y. & Rieh, S. Y. (2005). Web searching across languages: Preference and behavior of bilingual academic users in Korea. *Library & Information Science Research*, 27(2), 249–263.
- Rieh, S., & Xie, H. (2006). Analysis of multiple query reformulations on the web: the interactive information retrieval context. *Information Processing and Management*, 42, 751-768.
- Rieh, S. Y., & Xie, H. (2001). Patterns and sequences of multiple query reformulations in Web searching: A preliminary study. In E. Aversa, & C. Manley (Eds.). *Proceedings of the 64th Annual Meeting of the American Society for Information Science and Technology, vol. 38* (pp. 246–255). Medford, NJ: Information Today.
- Rho, J.-J. (2002). An investigation of the perceived information needs, information seeking behaviors, and the use of community public libraries among firstgeneration adult Korean immigrants living in the Dallas, Texas area. (Doctoral Dissertation, Texas Woman's University, 2002).
- Salton, G. (1973). Experiments in multi-lingual information retrieval. *Information Processing Letters*, 2(1), 6-11.
- Saracevic, T. (2007a). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915-1933.
- Saracevic, T. (2007b). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 2126-2144.
- Saracevic, T. (1999). Information Science. *Journal of the American Society for Information Science*, 50(12), 1051-1063.
- Saracevic, T. (1997). The stratified model of information retrieval interaction: Extension and applications. *Proceedings of the American Society for Information Science*, 34, 313-327.
- Saracevic, T. (1996). Relevance reconsidered. In P. Ingwersen & N. Ole Pors (Eds.) CoLIS2: 2nd International Conference on Conceptions of Library and Information science (pp. 201-18). Copenhagen, Denmark: Royal School of Librarianship.

- Saracevic, T. (1991). Information science: origin, evolution and relations. *In CoLIS 1. International Conference on Conceptions of Library and Information Science: Historical, Theoretical and Empirical Perspectives.* London: Taylor Graham.
- Saracevic, T. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society for Information Science*, *26*, 321–343.
- Saracevic, T., Kantor. P., Chamis, A. Y., & Trivison, D. (1988a). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39 (3), 161-176.
- Saracevic, T., & Kantor. P. (1988b). A study of information seeking and retrieving. II. Users, questions and effectiveness. *Journal of the American Society for Information Science*, 39 (3), 177-196.
- Saracevic, T., & Kantor. P. (1988c). A study of information seeking and retrieving. III. Searchers, searches and overlap. *Journal of the American Society for Information Science, 39 (3)*, 197-216.
- Serola, S. & Vakkari, P. (2005). The anticipated and assessed contribution of information types in references retrieved for preparing a research proposal. *Journal of the American Society for Information Science*, *55*(4), 373-381.
- Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*, 29, 3-48.
- Schamber, L. (1991). Users' criteria for evaluation in a multimedia environment. In J.-M. Griffiths (Ed.) Proceedings of the 54th Annual Meeting of the American Society for information Sciences(2), (pp. 126-133). Medford, NJ: Learned Information.
- Schamber, L., Eisenberg, M., & Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management*, 26(6), 755-776.
- Sewell, W., & Teitelbaum, S. (1986). Observations of end-user online searching behavior over eleven years. *Journal of the American Society for Information Science*, 37, 234-245.
- Shannon, C.E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Shaker, A. K. (2002). Bibliographic access to non-Roman scripts in library OPACs: A Study of selected ARL academic libraries in the United States. (Doctoral Dissertation, Pittsburgh University, 2002).
- Shaw, D. (1995). Bibliographic database searching by graduate students in language and literature: Search strategies, system interfaces, and relevance judgments. *Library* & Information Science Research, 17, 327-345
- Shera, J.H. (1970). *Sociological Foundations of Librarianship*. Bombay: Asia Publishing House.
- Shera, J.H. (1965). An epistemological foundation for library science. Cleveland, Press of Western Reserve University.
- Sheridan, P. & Ballerini, J. (1996). Experiments in multilingual information retrieval using the SPIDER system. *ACM*, 58-65.
- Shin, H.-S. (2003). Quality of Korean Cataloging Records in Shared Databases. Cataloging & Classification Quarterly, 36(1), 55-90.

- Slon, D. J. (2001). *Road maneuvers: Information seeking behaviors of public library users on the internet*. (Doctoral Dissertation, University of North Carolina, 2001).
- Spink, A., & Jansen, B. J. (2004). Web search: Public searching of the Web. Boston: Kluwer Academic Publishers.
- Spink, A. (2002). Multitasking information seeking and searching processes. *Journal of the American Society for Information Science*, *53*(8), 639-652.
- Spink, A., & Saracevic, T. (1993). Dynamics of search term selection during mediated online searching. In S. Bonzi (Ed.), ASIS '93: Proceedings of the American Society for Information Science (pp. 63-72). Silver Spring, MD: American Society for Information Science.
- Storer, N. (1970). The internationality of science and the nationality of scientists. *International Social Science Journal*, 22(1), 80-93.
- Su, S., & Conaway, C. (1995). Information and a forgotten minority: elderly Chinese Immigrants. *Library & Information Science Research*(17), 69-86.
- Tagliacozzo, R., & Kochen, M. (1970). Information-seeking behavior of catalog users. *Information Storage and Retrieval*, 26, 230-249.
- Tague-Sutcliffe, J. (1992). The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management*, 28(4), 467-490.
- Taylor, A. G. (2000). *Wynar's Introduction to Cataloging and Classification*. (9th Eds.). Englewood, Colorado: Libraries Unlimited, Inc.
- Taylor, A. G. (1984). Authority files in online catalogs: An investigation of their value. *Cataloging & Classification Quarterly*, 4(3), 1-17.
- Taylor, R. S. (1991). Information Use Environments. In Brenda Derwin & Melvin J. Voigt (Eds.). *Progress in Communication Sciences*. Norwood, NJ: Alex. (Vol. 10: pp.217-255).
- Taylor, R. S. (1962). The process of asking questions. *American Documentation*, *13*(4), 391-396.
- Taylor, R.S. (1968). Questioning-negotiation and information seeking in libraries. *College & Research Libraries*, 29, 178-194.
- Thomas, D.H. (1997). *The effect of Interface design on its, selection in an online catalog.* (Doctoral Dissertation, University Pittsburgh, 1997).
- Trzebiatowski, E. (1984). End-user study on BRS/After Dark. RQ, 23, 446-450.
- Vakkari, P. (2003). Task-based information searching. *Annual Review of Information Science and Technology*, 37, 413-464.
- Vakkari, P. (2001). A theory of task-based information retrieval. *Journal of Documentation*, 57 (1), 44-60.
- Vakkari, P. (2000). Relevance and contributing information types of searched documents in task performance. *Proceedings of SIGIR 2000 conference, Athens*, 2000 (pp. 2-9). New York: ACM.
- Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal case study. *Information Processing & Management*, 39(3), 445-463.
- Vilar, P. & Zummer, M. (2005). Comparison and evaluation of the use interfaces of ejournals. *Journal of Documentation*, 61(2), 203-227.
- Wang, P., Berry, M., & Yang, Y. (2003). Mining longitudinal Web queries: Trends and patterns. Journal of the American Society for Information Science and Technology,
54, 743–758.

- Wang, P., Hawk, W. B., & Tenopir, C. (2000). Users' interaction with World Wide Web resources: An exploratory study using a holistic approach. *Information Processing* and Management, 36, 229–251.
- Wertsch, J. V. (1985a). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (1985b). Culture, communication, and cognition: Vygotskian perspectives. Cambridge, MA: Cambridge University Press.
- Wertsch, J. V. (1991). Voices of the Mind: A Sociocultural Approach to Mediated Action. Cambridge, MA: Harvard University Press.
- Whitney, G. (1990). *Language distribution in databases: an analysis and evaluation*. Metuchen, NJ: The Scarecrow Press, Inc.
- Wildemuth, B.M. (2006). Evidence-Based Practice in Search Interface Design. *Journal of the American Society for Information Science*, 57(6), 825-828.
- Wildemuth, B.M., & Moore, M.E. (1995). End-user search behaviors and their relationship to search effectiveness. *Bulletin of the Medical Library Association*, 83, 294-304.
- Wilson, T.D. (1999). Models in information behaviour research. *Journal of n Documentation*, *55*(*3*), 249-270.
- Wilson, T.D., Ford, N.J., Ellis, D., Foster, A.E., & Spink, A. (2000, August). Uncertainty and its correlates. Paper presented at *Information Seeking in Context*, 2000, Gothenburg, Sweden.
- Xie, H. (2002). Patterns between interactive intentions and information-seeking strategies Source. *Information Processing and Management: an International Journal*, 38(1), 55-77.
- Yuan, W. (1997). End-user searching behavior in information retrieval: a longitudinal study. *Journal of the American Society for Information Science*, 48(3), 218-234.
- Zeng, L (1991). Automation of Bibliographic Control for Chinese Materials in the United States. *International Library Review*, 23, 299-319.
- Zeng, L. (1992). An Evaluation of the Quality of Chinese-Language Records in the OCLC OLUC Database and the Study of a Rule-Based Data Validation System for Online Chinese Cataloging. (Ph.D. dissertation, University of Pittsburgh).
- Zhang, F. & Zeng, M. (1998). Multiscript information processing on crossroads: Demands for shifting from diverse character code sets to the Unicode tm Standard in library applications. In 64th IFLA General Conference August 16-August 21, 1998. Retrieved on December 20, 2007 from http://www.ifla.org/IVifla64/058-86e.htm

CURRICULUM VITA

Yoo Jin Ha

EDUCATION	
May 2008	Ph.D. in Library & Information Scienc
	School of Communication, Information & Library Studies,
	Rutgers, The State University of New Jersey, New Brunswick.
February 2003	Master of Library Science,
	University at Buffalo, The State University of New York,
February 1991	Bachelor of Art in Library and Information Science,
	Yonsei University, Seoul, Korea

PROFESSIONAL EXPERIENCE

January 2006 – April 2007	Reference Librarian in the Information Services Department at the East Brunswick Public Library.
July 2005 – December 2005	Assistant in Inter-Library Loan operations at Rutgers University Libraries.
October 2003 – July 2005	Data Management Assistant and Digital Imaging Assistant Scholarly Commutation Center (SCC) in Rutgers University Libraries.
October 1990 - April 1991	Representative at COWBOY Korea, Seoul, Korea (an international subsidiary corporation)
April 1997 - July 2001	Director at JINHO Corporation, Seoul, Korea.
April 1991- April 1997	Senior Assistant at ITOCHU Corporation, Seoul Branch, Korea. Department of Provisions & General Merchandise

PAPER (REFEREED) / POSTER PRESENTATION

Ha, Y. (2008, January). *Implications of User Issues When Gaining Access to Non-English Information*. Poster presented at Doctoral Student Research Poster of the Association for Library and Information Science Education Conference, Philadelphia, PA. Ha, Y. (2006, June). Users' Information Seeking Behaviors With Cross Language Information Retrieval Systems. Paper presented at the Library Research Roundtable session in the 2006 American Library Association, New Orleans, LA

Ha, Y. & O'Connor, D. (2005, September). A Global Reach for the Information Disciplines. Paper presented at *i*-Conference, Penn State University.

Ha, Y. (2005, January). *Beyond Language Barriers: Survey of current LIS Education and Online library system*. Paper presented at the Association for Library and Information Science Education Conference, Boston, MA.

Ha, Y. (2005, January). *Beyond Language Barriers: Survey of current LIS Education and Online library system*. Poster presented at the Association for Library and Information Science Education Conference, Boston, MA.

Served as a paper reviewer for i-Conference, August 2005.