BIBLIOGRAPHIC ACCESS TO NON-ROMAN SCRIPTS IN LIBRARY OPACS: A STUDY OF SELECTED ARL ACADEMIC LIBRARIES IN THE UNITED STATES

By

Ali Kamal Shaker

BA—Cairo University-Egypt, 1995

MLIS—University of Wisconsin-Milwaukee, 2000

Submitted to the Graduate Faculty of

School of Information Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2002

DISSERTATION DEFENSE

Name of Student: Ali Shaker

Title: Bibliographic Access to Non-Roman Scripts in Library OPACs: A Study of Selected ARL Libraries in the United States

Committee:

Name	Accept	Department
Arlene Taylor	X	DLIS
Edie Rasmussen	X	DLIS
Brian Butler	X	Katz Graduate School of Business
Rush G. Miller	X	Hillman University Librarian & Director

Major Advisor Arlene G. Taylor
Arlene Taylor

Date 12/18/02

BIBLIOGRAPHIC ACCESS TO NON-ROMAN SCRIPTS IN LIBRARY OPACS
A STUDY OF SELECTED ARL ACADEMIC LIBRARIES IN THE UNITED STATES

Ali K. Shaker, PhD

University of Pittsburgh, 2002

With the increasing availability of non-Roman script materials in academic/research libraries, bibliographic access to vernacular characters in library OPACs becomes one of the primary means for users of these materials to access and use them efficiently. Though Romanization, as a bibliographic control tool, has been studied extensively during the past five decades, investigations of using original (vernacular) scripts remain inadequate.

The purpose of this study was to trace the transition from Romanization-based to vernacular-based bibliographic access to non-Roman script materials. Two major developments contributing to this transition were the availability of records with script characters from bibliographic utilities, and the development of a universal character set, the Unicode standard.

The main data collection instrument was a self-administered mail questionnaire sent to a purposive sample of academic library members in the Association of Research Libraries with sizeable non-Roman script collections. Another data collection technique utilized was document/Web site analysis of bibliographic utilities and library automation vendors. Forty five questionnaires were obtained, which represented 65% of an actual population of 69 libraries.

The major conclusions of this study are: (1) Most academic libraries catalog their non-Roman script materials using vernacular characters in bibliographic utilities, not on their OPACs; (2) Despite the advances in multilingual support capabilities in automated library systems, academic libraries are still unable to exploit these systems to their maximum benefit;

(3) The majority of libraries surveyed performed CJK cataloging in vernacular characters, but Cyrillic cataloging is still in romanized format; (4) CJK and Arabic/Hebrew librarians showed a strong attitude toward vernacular cataloging; Cyrillic librarians showed significant opposition (42%) in their attitudes toward Cyrillic vernacular cataloging; (5) Catalogers of CJK and Arabic/Hebrew materials experienced various difficulties with Romanization; Cyrillic catalogers had the least trouble with Romanization; and (6) Among the systems librarians surveyed, 58% indicated future plans considering scripts support.

In addition, the study has developed a set of functional requirements for OPAC designers and systems managers with regard to vernacular access to script characters in bibliographic records. The future research directions recommended concern the future of Romanization as a bibliographic control tool, end users' familiarity with transliteration tables, re-cataloging romanized records to include scripts, and multilingual authority files.

To the memory of my father, Kamal
who could not wait to celebrate my success

and

To my newborn son, Kamal
Who comes to life
near the defense of this dissertation

Acknowledgement

My deepest appreciation goes to Dr. Arlene Taylor, committee chair and advisor, for her generous support throughout this dissertation process. Her expertise, guidance, and understanding were of enormous help to me. It was my great fortune to work with my committee members Dr. Edie Rasmussen, Dr. Rush Miller, and Dr. Brian Butler who enthusiastically responded to the research. Each of the committee members played a special role that corresponds, in some part, to particular aspects of this study. I have total respect and admiration for the work they have done and that they continue to do.

I want to thank the librarians of the University of Pittsburgh for helping with pre-testing the questionnaire. I also want to acknowledge the assistance of Mr. Donald King, School of Information Sciences and Dr. Thomas Zullo, School of Education, for checking and advising on the statistical analyses in the study.

Last but not least, to my large family overseas, thanks for encouraging me till the completion of this dissertation, and I am sorry that you had to survive without me.

Finally, special thanks go to my wife, Eman, and lovely daughter, Sohayla. They have been so supportive and encouraging. Without them being with me here, it would be so difficult to get this done just on the right time.

Thank you all for all that you have done for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background

Presenting vernacular bibliographic data in library OPACs was a concern in most libraries that have sizeable collections of non-Roman language materials. Traditionally, libraries with non-Roman collections, especially those in North America and Western Europe, separated these materials from their main catalog and, thus, overcame the various filing and sorting problems. Other libraries applied techniques of Romanization and transcribed non-Roman scripts to the Roman alphabet. None of these procedures is an adequate way to solve the problems associated with location and access of non-Roman language materials.

The situation in general did not improve when OPACs were introduced. Very often non-Roman scripts proved to be difficult in data processing; so these materials were excluded, and library procedures and services for these collections remained unchanged by automation. However, during the past two decades, two major developments occurred that could facilitate the shift or the transition of processing non-Roman script materials in the vernacular. The first development is the work of bibliographic utilities, started in the early 1980s, in providing bibliographic records with original (vernacular) characters, and the second is the development of a universal character set, the Unicode standard in the early 1990s, that will overcome the technical difficulties of handling non-Roman scripts simultaneously with Roman scripts in the same application (e.g., automated library systems).

Professional organizations have identified the importance of vernacular bibliographic data as a natural way to access non-Roman language collections. "*Libraries should catalog all*

materials in the original language and script. They should provide subject access both in English and in the original language." (Guidelines for Multilingual Materials…, 1990)

The importance of vernacular bibliographic records in library OPACs is being increasingly recognized as a necessity for the effective use of these script materials. The availability of original scripts in library catalogs is not only valuable to librarians who must use transliteration schemes to artificially represent the bibliographic data for non-Roman script materials in Roman characters, with all the problems that may occur during this practice, but is also critical to effective inquiry of library OPACs to ensure that the end user will retrieve all relevant items. As we will see in chapter two, the literature review, many authors questioned the value of transliteration, considering such practice impractical, expensive, and frustrating to the end user who knows non-Roman scripts, as well as to the user who does not.

The purpose of this study is, therefore, to investigate the development of bibliographic access to original scripts in library OPACs. The intent is to trace the transition that the two major developments explained above may have invoked in handling non-Roman scripts in bibliographic records. This study arises from three questions: (1) Are current OPACs in academic libraries equipped with support for non-Roman scripts? (2) To what extent do academic libraries consider the inclusion of vernacular characters in their OPACs? And (3), what specifically is necessary to establish a non-Roman script-supported OPAC?

1.2 Statement of the Problem

The need for non-Roman scripts in library OPACs has been receiving more attention since the early 1980s when bibliographic records started to host original (vernacular) data in the two major bibliographic utilities, OCLC and RLIN. This situation has been emphasized by limitations placed by current OPACs on the dissemination of information within academic and research libraries that are detrimental to the effective and efficient access to non-Roman script materials. Current practices of Romanization hinder access to non-Roman script materials which is important to students and researchers in area and foreign languages' studies.

One of the most demanding responsibilities of librarians in many academic, research, and special libraries is dealing with non-Roman language materials. Better access to these materials is important to the sustained growth of academic and research libraries that hold unique and difficult-to-access collections of these materials. This requires providing access to the bibliographic records in the vernacular, where end users are able to search and display the original scripts in library catalogs.

Despite the prominence of English and other Roman languages in modern book publishing, many non-Roman languages are increasingly contributing to the published knowledge base of human beings. Many academic and research libraries in the United States acquire information resources in these languages to better serve the teaching and research needs in their institutions. In fact, the problems of processing non-Roman script materials are becoming more complicated with the growth of the publishing industry in Asia and the Middle East. Table 1.1 below shows some excerpted numbers for book production in some non-Roman language speaking countries around the world.

Table 1.1 Book Production: Number of Titles by Universal Decimal Classification Classes

Country	Year	Total
Egypt	1991	2 599
	1993	3 108
	1995	2 215
China	1990	73 923
	1993	92 972
	1994	100 951
Israel	1985	2 214
	1992	2 310
Japan	1992	35 496
	1996	56 221
Saudi Arabia	1980	218
	1996	3 900
Korea, Republic of	1994	34 204
	1995	35 864
	1996	30 487
Russian Federation	1994	30 390
	1995	33 623
	1996	36 237

(Source: *UNESCO Statistical Yearbook* 1999. Paris: UNESCO. p. IV-80)

This volume of published materials and the trend in most academic and research libraries to collect it necessitates that library OPACs should, as gateways to knowledge regardless of language, support non-Roman scripts to better handle the information needs of library users. Unfortunately, this is not the case nowadays. Most OPACs have been developed in an English language environment, and hence have complete support for English regardless of who is going to use them. However, most current OPACs in academic libraries do not support vernacular access to non-Roman scripts.

1.3 Need for the Study

The steady flow of literature written in non-Roman scripts and the growth of libraries serving multilingual user populations emphasize the growing importance of bibliographic access to original characters in library OPACs. The studies and surveys cited in chapter two have not been specific in analyzing and identifying bibliographic access to non-Roman scripts in library catalogs in academic libraries in the United States, nor have they examined whether these libraries changed their practices along with the transition in the developments in the bibliographic utilities and the technological advances in encoding non-Roman characters (e.g., Unicode). Furthermore, few studies have looked at functional requirements with regard to what sort of non-Roman script support features should be presented in library OPACs in order to fulfill the multilingual needs of academic libraries with sizeable non-Roman script materials.

Therefore, this study is intended to determine the transition into vernacular bibliographic access to non-Roman scripts in library OPACs, and also, to determine a set of functional requirements for current and future OPACs. While the former purpose aims to illustrate the current status of vernacular characters in bibliographic records in current OPACs, the latter is intended to provide recommendations on functional requirements in future OPACs to better support creation, searching, and displaying of original scripts.

1.4 Purpose of the Study

The purpose of this study is three-fold based on the following cornerstones: availability, future plans, and functional requirements. In more detail, it will:

1. Determine the "*availability"* of non-Roman scripts in current library OPACs in selected academic libraries in the United States. The intention here is to document how the

transition, spurred by developments in bibliographic utilities and characters encoding, affected bibliographic access to vernacular characters in library OPACs.

2. Gather as much information as possible about academic libraries' "*future plans"* for the inclusion of non-Roman scripts in their OPACs. The intention here is to determine how libraries consider bibliographic access will be given to original scripts in the near future (within 5 years).
3. Provide recommendations to designers and developers of library OPACs, and to librarians as well, with regard to "*functional requirements"* toward supporting the creation, searching, and displaying of vernacular characters in library catalogs. The intention here is to establish some features that could guide the decision-making process in automation vendors' upgrading plans, as well as to guide academic libraries when communicating with vendors.

1.5 Definition of Terms

Script. According to the *Oxford English Dictionary*, "script" could be defined as, "A kind of writing, a system of alphabetical or other written characters." This is different than **language**, that means, "The whole body of words and of methods of combination of words used by a nation, people, or race; a `tongue'." An example is the Arabic script that is used as the writing system for different languages, such as Arabic, Persian, and Urdu.

Vernacular. The OED lists many uses for the word vernacular. Of particular interest to this study are two usages. It could be used as an adjective, "*That writes, uses, or speaks the native or indigenous language of a country or district*." And it could also be used as a noun, "*The native speech or language of a particular country or district*."

Non-Roman collections. Library collections that are written in scripts other than the Roman (or Latin) scripts. These collections include, but are not limited to, materials in East Asian (e.g., Chinese, Japanese, Korean), right-to-left (e.g., Arabic, Hebrew), and Cyrillic (e.g., Russian) scripts.

Interface language. The interface language of the automated library system allows the user to choose the language of menus, dialog boxes, help screens, etc., that is desired. However, this does not mean changing the language of the keyboard. For example, a user can choose Russian as the interface language while still typing in English.

1.6 Overview

This study was designed to trace the transition in bibliographic access to non-Roman scripts in library OPACs, and the role of bibliographic utilities and automated library systems in this transition. This introductory chapter has presented background information that is necessary for this investigation.

Chapter two provides a review of the literature that is relevant to different sub-categories that make up the issue of vernacular access to non-Roman characters in bibliographic records, and the prevailing forces that impact the transition to such access.

Chapter three describes the research design, survey instrument, sample frame, procedures to pre-testing the questionnaire, and research procedures that were utilized to collect and analyze the data for the study.

Chapters four and five present the data analyses and conclusions of the study, respectively. Chapter four reports the results of the statistical analyses that were used to describe the sample and address the research questions. Chapter five presents a summary of the study's findings, discussions, implications, and recommendations for future research.

CHAPTER TWO

LITERATURE REVIEW

Since the wide introduction of automated library systems in all types of library and information services and operations, a series of studies have been conducted to explore issues concerning bibliographic access to non-Roman scripts in library OPACs. Recent developments have seen a gradual increase in the number of non-Roman bibliographic records available through the bibliographic utilities. This has resulted in an increased pressure on academic and research libraries to enhance bibliographic access to these materials through vernacular searching and display in their OPACs.

The practice of Romanization as the basis for cataloging non-Roman language materials, though it may have originally been the only solution because of software limitations, needs to be looked at again in the light of new developments in vernacular bibliographic descriptions; and there are, in fact, indications that it is being reconsidered. However, improvements in vernacular bibliographic descriptions alone are of limited value if the designers of automated library systems and librarians are not aware of the vital importance, to the user of bibliographic records, of their being in the original script.

There are some hopeful signs that the situation is changing. Several authors have discussed the problems for users caused by Romanization, and recent developments in computer software and standards will eventually provide ways to overcome the current limitations of handling non-Roman scripts. (Leggott, 1991; Petzold, 1993) These standards were incorporated into USMARC and UNIMARC (USMARC; Holt). Today, there is a new universal multiscript

character set, Unicode, which is being implemented in products from leading computer companies (Unicode Consortium).

This review is divided into three sections. The first section provides a synopsis of some of the "*general works*" in this area. The second section, "*main themes,*" illustrates an overview of the conceptual and theoretical framework that supports the study. And finally, the third section examines the literature reported as "*empirical research*" on the problem.

General Works

As early as 1934, Sommer (1934), in a letter to the *Library Journal*, suggested that readers "*might be able to help themselves to a greater extent if they had a special index of at least the titles in the native script and in the order of their alphabet.*" Starting in the 1970s, discussions began on the disadvantages of Romanization of bibliographic records, and examination of possibilities and advantages of presenting non-Roman scripts in library catalogs has received increasing attention. In the words of one of the chief advocates of the use of non-Roman scripts, Hans Wellisch (1976), "*one must ask whether the practice* [of transliteration] *as performed by the majority of libraries is really the only reliable and "scientific" bibliographic control tool it is claimed to be by librarians and bibliographers, or whether their goals and those of library users alike could perhaps be achieved more effectively by other means.*"

The International Federation of Library Associations and Institutions' (IFLA) interest in the problems of multilingual and multi-script issues in automated library systems extends well into the past. More than a decade ago, in fact, the first IFLA pre-conference on this topic was held in Tokyo and resulted in publication in 1987 of "*Automated Systems for Access to Multilingual and Multiscript Library Materials—Problems and Solutions*." (Bossmeyer, 1987) Although the conference emphasized oriental scripts since it was held in the Pacific region, the

issues identified were generally applicable to other non-roman scripts as well. At the conclusion of the pre-conference, Durance (1987) summarized several points that had emerged during the conference deliberations as topics requiring priority attention, including:

- The need for more influence on technical systems development and participation in standards development from countries and areas for which various languages and scripts are in the vernacular.
- Need for standards for sorting and retrieval of all scripts, but in particular ideographical scripts.
- Adjustments of the ISBDs and of UNIMARC to better accommodate non-roman scripts.
- Discovery of incentives to persuade vendors to develop software and hardware to better support multi-script applications.

As a follow-up to this session, the IFLA Sections on Information Technology and on Library Services to Multicultural Populations joined with the IFLA Division on Bibliographic Control in sponsoring a satellite meeting in August 1993 just prior to the annual IFLA conference in Madrid which dealt with issues related to automated systems for access to multilingual and multi-script library materials. This meeting was designed to respond to increasing international interest and concern over potential growing and divergent technological responses to display and access issues for multilingual and multi-script library materials. The satellite meeting's official goal was to "*focus on multilingual and multi-script problems in the library service arena, especially in the automated environment and to facilitate the access to this sort of material*" (McCallum, 1993).

The 1993 program covered such topics as the design of multilingual and multi-script catalogs, multi-script issues and graphic displays for Online Public Access Catalogs (OPACs), and automation of various scripts within online systems.

The satellite meeting covered resolutions and action items which were approved by the IFLA Section on Information Technology and the IFLA Section on Library Resources to Multilingual Populations. The IFLA Section on Cataloging adopted four of the action items as part of its action plan for its Medium Term Programme:

- Encourage changes to cataloging rules to include all scripts and languages;
- Develop recommendations on sorting arrangements across languages and scripts;
- Develop recommendations for treatment of word division in appropriate non-roman scripts; and
- Develop recommendations for treatment of abbreviations in bibliographic records to reduce the difficulty of comprehension of data in multilingual databases.

The Cataloging Section's 1992-1997 Medium Term Programme had already featured two relevant goals, which provide the essential framework to these action items.

- Propose methods to improve the handling of multi-script records and the integration of such records in various types of bibliographic databases.
- Assess the problems and prospects of linking various single language and/or multilingual name authority files.

As part of its response to these action items, the Section on Cataloguing's Standing Committee encouraged the work of Murtomaa and Grieg who in 1994 published an article in

International Cataloguing and Bibliographic Control that discussed the problems and prospects of linking various single-language and/or multi-language name authority files. This article was based on discussions held from 1991 to 1993 within meetings of the IFLA Standing Committee for the Section on Cataloging.

The Section's Standing Committee also followed up by arranging, as a featured part of its open program at the 61st IFLA Conference in 1995, for John Eilts to present a paper on how North American libraries are dealing with non-roman script materials for automation and international exchange. Eilts (1996) concluded his paper with the recommendation to begin developing necessary programs to translate relevant standards (e.g., character sets, data formats, subject classification, etc.) "*to and from those in use in countries where scripts are used, and in areas where the direct translation of formats would cause a loss of data*."

As a further contribution to consideration of issues raised at the Satellite Meeting held at Madrid, the Section on Cataloguing sponsored a workshop in Istanbul, Turkey, August 24, 1995 (Byrum, 1998). At that gathering, six participants presented major papers covering key issues.

While the authors dealt with a variety of different national and international issues, several recurring themes emerged from their papers. The themes involved the increasing need for standards and cooperative sharing of bibliographic and authority data as the only viable methods to solve the seemingly impossible problems associated with multi-script bibliographic environments. The authors suggested ways of solving some of the issues related to display and access on national bases that may have international implications. Other authors voiced the undeniable need for expanded standards, such as Unicode, that offer the potential to standardize a character set that could cover all scripts. Above all, the role of national bibliographic agencies is seen as even more critical than ever in both national and international arenas.

While the authors of the papers suggested many more specific recommendations related to the individual topics, they agreed on several over-arching themes. In many cases the authors looked to IFLA, national bibliographic agencies, developers of standards such as ISO and MARC formats, and the commercial library automation industry to respond to these recommendations.

There is a broad responsibility for pursuing the following recommendations:

Bibliographic and Authority Files

- Include as essential components of our national and international bibliographic databases the abilities to include and display records with original language scripts. These databases should provide links between different forms of records and should include and display original language scripts and links between different forms of language and scripts of name and subject headings.
- Jointly create and maintain multilingual and multi-script authority files to be used nationally and internationally. Within these authority files, recognition should be given to different authorized forms of names.

Role of Bibliographic National Agencies/IFLA/ISO

- Determine on national and international levels the roles that centralized agencies, such as national bibliographic agencies, should play in the development of shared bibliographic and authority data files (including the verification of uniform headings).
- IFLA and ISO have crucial and continuing roles in the development and promotion of international standards.

Importance of Standards

- Promote national and international creation/use of standards and problem solving related to issues of access, display and maintenance of bibliographic and authority data files.
- Create a universal character set that would cover all major scripts and promote its general international use.
- Develop an agreed-upon mapping between current character sets and the new universal character set.
- Develop a default sorting order for the full character repertoire of ISO 10646[*].
- Standardize "filing rules" in multilingual and multi-script environments.
- Create standards to facilitate resource discovery on the Internet and the use of mark-up languages such as SGML and HTML.
- Reconsider the MARC format and current cataloging codes and their applicability to Unicode and multi-character set conventions and requirements.

Education and Research

- Promote within individual countries and regions cooperative bibliographic efforts and the importance of bibliographic standards, while recognizing and balancing legitimate needs for local practices.
- Increase awareness of how technology can be used for resource sharing and cooperative bibliographic control.
- Increase emphasis on the need to search and display original scripts.

[*] ISO 10646, published in 1993, is the first officially standardized coded character set with the purpose to eventually include all characters used in all the written languages in the world. As the relation with Unicode character set, Unicode is kept, from version 1.1, compatible with ISO 10646. The Unicode Consortium, that is constituted of major American computer manufacturers, is also an important contributor to the ISO work to further develop ISO 10646.

- Support additional research on sophisticated search engines that can cope with multiple scripts and languages and improve recall for natural language queries. In particular continue solving problems associated with the innate differences among major international scripts.

International organizations were strongly interested in the problems of handling multilingual and multiscript materials in the automation environment. Many librarians from the Western world also explored the dilemma of cataloging works in writing systems other than the Roman alphabet. Agenbroad (1992) reviewed some characteristics of these writing systems, and the implications of these characteristics for input, retrieval, sorting, and display needed for adequate online catalogs of such works. The following are four groups into which non-Roman scripts are generally divided for simplicity and features that have implications for cataloging: (1) European scripts—upper and lower case (Greek, Cyrillic, and Armenian); (2) Semitic scripts—read right to left (Hebrew and Arabic); (3) Indic scripts—implicit vowel (indigenous scripts of India and Nepal); and (4) East Asian Scripts—very large character repertoires (Chinese, Japanese, and Korean). Agenbroad concluded that "the basic problem will soon be political, not technical."

Main Themes

This section of the review is presented in five main themes. The first theme deals with "*transliteration,*" the dominant method for presenting non-Roman scripts in library OPACs, with special emphasis on its problems and solutions to overcome them. The second theme describes "*organizations*" concerned with vernacular bibliographic records (e.g., bibliographic utilities,

libraries). The third theme examines emerging "*tools*" in multilingual access to library OPACs, including communication formats (e.g., MARC, UNIMARC), multilingual authority files, character sets (e.g., Unicode), and multilingual interfaces. The fourth theme briefly outlines "*rules*" dealing with multilingual bibliographic information in cataloging codes (e.g., AACR2, ISBDs). The fifth theme provides an overview of some "*examples*" of bi- and multi-lingual automated library systems, mostly in countries where non-Roman scripts are used. Finally, a "summary and conclusion" ties the work together by recapitulating the themes presented in the review. The chapter ends with a special section that illustrates some examples of the "*empirical research*" that has been conducted on the topic.

I. Transliteration

Transliteration can be defined as the representation of the characters of one language, whether they are ideographic (such as CJK), or alphabetic (such as Arabic, Hebrew, Russian, etc.) in the orthography of another, usually in order to facilitate their integration into the dominant alphabet of the cataloging agency. This distinguishes transliteration from *transcription*, which deals with the sounds of language rather that its written characters. As stated by Weinberg (1974), one could transcribe but never transliterate Chinese. One could transcribe or transliterate Yiddish, but each process would yield a different result.

Romanization is a specific term meaning the conversion of names or text not written in the Roman alphabet into Roman alphabet form (AACR2R, 1998). *Reversibility* is a feature of exact transliteration only, because each letter of the foreign alphabet is assigned a distinct corresponding symbol in the Roman alphabet, and conversion in either direction is a simple matter of table look-up. In an ideal world, transliteration and transcription would be the same and

completely reversible (Weinberg, 1974). However, Blanken (1971) stated that, "*perfect reversibility is difficult to achieve in practice without some knowledge of the source language*."

In the literature, studies proliferate that discuss the process of transliteration and its intricacies. While many of these studies have focused on the "how" of transliteration, few have concentrated on the "why." This situation provoked some authors to speculate on some possible reasons. Weinberg (1974) pointed out that a first reason is the dominant idea of the "integrated catalog," where all catalog entries are in one alphabet. And second, a more plausible reason for Romanization is our basic tenet, 'the purpose of a catalog is to keep all the works of an author together.' But, can't librarians create cross-references between the romanized and the vernacular script entries? And, why do librarians transliterate title entries? Weinberg also tried to find a reason for that, "*to allow the typing and filing of these entries to be done by the regular clerical staff*." But the money saved by not hiring clerical staff with knowledge of foreign alphabets is completely spent on the Romanization process, cross-referencing from alternate forms, and we still have no guarantee that the user will find the item he/she seeks. Even Romanization of author names cannot be justified economically. An enormous number of cross-references must be made from various popular and scholarly forms.

If this is so, then the next question, as asked by Spalding (1977), must be: "*Why romanize for filers if this inconveniences readers?*" How do we answer this except to say, "*Because any other course would be contrary to our concept of the universal catalog?*" Now back to the idea of the integrated catalog again. This seems to be the main reason for cataloging non-Roman alphabet materials by means of transliteration: a catalog in which all items in the collection are entered in a single alphabet from A to Z, regardless of language, regardless of form, regardless of

subject. That was much the situation in the manual (e.g., card catalogs) environment. The online catalog environment has made these reasons, specially the last one, belong to the past.

Problems of Transliteration

After briefly discussing the linguistic functions and the correlation between the phoneme and the grapheme, Weinberg (1974) illustrated how the Roman alphabet is poorly equipped to represent the thirty-two primary phonemes or forty-three sounds of the English language. Then she asked, "*Are we going to burden these twenty-six letters with the representation of the sounds of all the languages of the world?*"

The multiplicity of scripts, unequal number of characters in different scripts and the inability of the characters of an alphabet to represent the sounds of different languages are of concern not only to linguistics but also to librarians. The Roman alphabet has 26 characters, while these vary from 30 to 40 in Cyrillic. Chinese has 1800 everyday life characters. Another problem of the scripts is that with the change of time, some of the letters are dropped and new characters are added to some scripts (Chakraborty, 1976).

Wellisch (1978) presented these problems in two categories, as follows:

(1) The Difficulties of Romanization for Operators (e.g., catalogers)

Librarians and bibliographers, that is to say, people who actually apply the method as operators of their bibliographic control systems, are confronted with the following problems:

a. *Multiplicity of Schemes*
b. *Inconsistent application and local adaptations*
c. *Variant forms of romanized names*
d. *Reversibility*
e. *Susceptibility to errors*
f. *Alphabetization*[*]

[*] Refers to the process of interfiling entries for documents originally written in Roman script with those that are artificially romanized which results in many inconsistencies.

(2) The Difficulties of Romanization for Users

Since library catalogs are designed specifically to facilitate access to library collections by end users, a look at problems encountered by these users is inevitable. In addition to all the problems discussed before, users are faced with an almost complete lack of keys to the Romanization system in use, which is almost never made explicit to users of catalogs or bibliographies. The burden of deciphering a Romanization in order to arrive at the original form of a name is entirely on the user and it amounts in many instances to exercises in cryptography, or the decoding of a message to which the code key is available only to the senders (the librarians) but not to the receivers (library users).

Weinberg (1974) pointed out that, "*the more we think of users, the less rationale we find for transliteration*." If people take the trouble to learn another alphabet, why should they have to learn an additional artificial system to get access to material in that alphabet?

Solutions

A quarter century ago, Spalding (1977) stated that, "*my purpose here is … to reexamine the basis of our present practice with respect to Romanization in an attempt to convince you that we have been on the wrong track all along and that a radical change is needed*."

By 1977, following a considerable discussion of the problem at the IFLA Worldwide Seminar in Seoul the previous year, the International Congress on National Bibliographies, organized by IFLA and held under UNESCO auspices in Paris, recommended that, "*wherever possible these records should be in the language and/or scripts in which the publication originally appeared*" (Simsova, 1985)

Describing transliteration as a "bibliographic dilemma," Bachman (1989) pointed out that, "*The bibliographic field badly needs a solution to this problem of conversion*."

Some possible alternative solutions have been suggested by Wellisch (1978):

(a) Separate listings by script and language

(b) Problems of logographic scripts[*]

(c) Centralized collections

(d) Unique identification of authors' names

(e) Unique identification of documents

A different set of solutions to the problems of transliteration has been given in a different paper by Wellisch (1980):

(a) Catalogs and bibliographies should clearly state which Romanization schemes they adhere to, including local modifications or exceptions; the key to the scheme should be displayed prominently to enable users to understand how names and words have been romanized, and (if such is possible) to enable them to reverse the Romanization to its original form.

(b) The titles of works in non-Roman scripts should not only be romanized but also translated into the language predominant among the users.

(c) Library catalogs and most printed bibliographies with romanized entries should have a counterpart of entries by parallel and/or complementary listings of documents in their original script with cross-reference from the romanized entries to the relevant entries in the original script.

[*] For example, the use of numerical codes to represent Chinese logograms.

More recent solutions are presented by Erickson (1997):

1. Character Images: This solution involves the use of Graphics Interchange Format (GIF), Portable Network Graphics (PNG), or other images to display the characters that are not available in the local character set, where the user is not likely to search for that particular piece of text.
2. Font-Related Solutions: Before moving on to font-related solutions, a distinction needs to be drawn between fonts and character sets. Briefly, Fonts contain glyphs, not characters. The code positions of the glyphs are not important, but the positions of characters in a character set are. Many companies produce fonts for Windows and for Macintosh with non-Roman characters placed in the positions usually reserved for Roman characters. In these fonts, what appears when you type an "a" might be a Greek alpha or Cyrillic R. This circumvents the need to use escape codes to switch character set encoding for non-Roman texts. With these fonts in hand, it is a simple task to write a multilingual word-processed document—simply change the local font when you want other glyphs. The initial problem with this is that you cannot use it on another computer of the same type unless it has the same non-Roman fonts installed and available.
3. Character Set Switching: Unambiguous communication of text is much simpler when some recognized standard is followed within a document. Mapping tables can be used to convert from one character set to another, assuming no changes have been made to the source character set locally. This is especially useful in an international environment, such as the Web, where no assumption can be made that others are using the same character set that you are using.

4. Unicode: In order to overcome the problems inherent in many character sets, Joseph Becker of Xerox PARC, along with Lee Collins and Mark Davis of Apple Computer, began work on a new character encoding. Becker named it "Unicode" for its aim to embrace three characteristics: universality, covering all modern written languages; uniqueness, with no duplication of characters even if they appear in more than one language; and uniformity, with each character being the same length in bits. The Unicode Consortium was incorporated in 1991 as Unicode, Inc., with members including the Research Libraries Group, Metaphor Computer Systems, Microsoft, IBM, Sun Microsystems, DEC, Adobe, Claris, NeXT, Pacific Rim Connections, Aldus, Go Corp., Lotus, Word Perfect, and Novell. Unicode is defined as "*a world-wide character encoding standard based on a 16-bit unit of encoding developed by Unicode, Inc*."

The present and potential possibilities of computers ought to be harnessed to the task of recording, storing, and retrieving documents irrespective of the graphic form in which they are written, and in a manner that makes it possible for those who can read them to do so without having to overcome artificial barriers (e.g., Romanization) to access. It will not be easy to do this, not so much because of the technical difficulties involved, but because long standing traditions, vested interests, and even rank prejudice will have to be fought (Wellisch, 1978).

The bibliographic control of the world's intellectual achievements should become truly universal. It should not be limited to the Western world nor geared only to the Roman script as the sole key to access, but should encompass records in all languages and scripts.

II. Organizations

Non-Roman scripts in Bibliographic Utilities (OCLC, RLIN)

A bibliographic utility is "*an organization which maintains online bibliographic databases, enabling it to offer computer-based support to any interested user. It provides a standard interface by which bibliographic records are available to libraries*" (Young, 1983). In the United States, there are two major bibliographic utilities serving not only American libraries, but international ones as well. These two are the Online Computer Library Center (OCLC), founded in 1967, and the Research Libraries Information Network (RLIN), developed by Research Libraries Group (RLG) in 1978.

In September 1983, RLG installed enhancements to RLIN to enable users to create, store, retrieve, display and transmit bibliographic records containing Chinese, Japanese, and Korean (CJK) vernacular data. To communicate such data successfully, two changes were designed for the USMARC format: a "character sets present" field to identify alternate character sets, and a new set of fields to hold the "alternate graphic representation" of romanized fields within a record (Bales, 1988). All Chinese, Japanese, and Korean characters are stored and transmitted using the 3-byte RLIN East Asian Character Code (REACC).

In April 1985, RLIN was enhanced to support the processing of information represented in the Cyrillic alphabet, allowing the use of Russian, Belarusian, Bulgarian, Macedonian, Serbian and Ukrainian. The character repertoire consists of the basic Cyrillic character set and the extension of the Cyrillic alphabet for bibliographic use adopted by ISO (Bales, 1988). Software was developed to allow the IBM PC to emulate RLG video terminals, process Cyrillic and Roman alphabet materials and to search the entire set of RLIN files. As in East Asian vernacular support, Cyrillic records also carry Roman and non-Roman data in parallel fields. The Cyrillic

characters are integrated into the RLIN indexes. RLIN followed by providing vernacular support to Hebrew and Yiddish in 1988, and Arabic in 1991, with the same capabilities as for CJK and for Cyrillic scripts.

Almost immediately after the successful completion of the RLIN CJK system, OCLC announced its plan to develop a similar system that would process CJK vernacular scripts. An OCLC press release of October 18, 1983, reported that OCLC and Asiagraphics of Mt. Sinai, New York, had entered into negotiations leading toward the development of a CJK library support package (Lee, 1988). In March 1985, OCLC invited a host of interested East Asian librarians to a CJK Advisory Meeting at its headquarters in Dublin, Ohio. Asiagraphics was dropped; a new partner, Eastern Computers, Inc., was chosen; and the package, formerly named the OCLC CJK350 system*, was completed for field test in May 1986. It allows the user libraries to create, edit, and retrieve bibliographic records containing CJK vernacular characters. An earlier agreement with RLIN has made it possible for OCLC to use the same character code, the REACC, and its character set for machine representation of the CJK vernacular scripts. The use of the same code will facilitate future record exchange among libraries in different bibliographic networks. In 2001, OCLC introduced Arabic cataloging software that allows cataloging Arabic language materials using Arabic script.

Zeng (1991) compared RLIN CJK and OCLC CJK systems, and concluded that the RLIN system provides a more advanced search structure and an online CJK thesaurus. Since it has a larger CJK database than OCLC, searching and evaluation of the matching records on RLIN are less time-consuming than on OCLC. According to the experience of the University of Illinois regarding the cost of operating a CJK system, RLIN is more expensive in both hardware and maintenance (Wei, 1989). On the other hand, since the OCLC CJK system was developed a few

* Now called, OCLC CJK software

years later than the RLIN system, it was able to take advantage of more advanced technology and offer some "state of the art" features. For instance, it uses a standard OCLC English language keyboard rather than one that is specially designed. For generating CJK characters, it provides multiple input methods. Furthermore, the OCLC CJK system is PC-based and offers three software packages for online cataloging, card production and word processing.

The quality of, and commonly occurring errors in, the member-contributed Chinese records in a large multilingual bibliographic database, the OCLC database, was the focus of Zeng's dissertation in 1992. In order to develop a viable approach to improve the quality of these bibliographic records, she developed a set of production rules for a data validation system for Chinese online cataloging that could be applied to bilingual bibliographic records containing vernacular characters. Instead of adding new and additional responsibilities for the catalogers, the proposed system would have the goal of reducing human efforts devoted to full-level cataloging and increasing productivity.

In addition to CJK, bibliographic control and retrieval using vernacular characters are also available for titles in Arabic, Persian (Farsi), Urdu, Ottoman Turkish, and other languages written with Arabic script in RLIN. Arabic script is the most recent addition to the scripts available on RLIN. Aliprand (1992a) provides a comprehensive description of the Arabic script features of RLIN. He covers Arabic character sets and RLIN's character repertoire for Arabic script; how Arabic characters are input and stored in the RLIN database; the equipment needed for Arabic script support; the indexing, retrieval, and representation of records containing Arabic script; the inclusion of non-Roman data in USMARC bibliographic records; and statistics on the RLIN database.

The addition of Arabic script capability to RLIN is the culmination of a decade-long effort to provide automated support for the bibliographic control of materials written in non-Roman scripts. The use of original Arabic script for retrieval instead of unnatural Roman characters will greatly facilitate library service to readers of languages written in Arabic script. The support Research Libraries Group (RLG) received from the Kuwait Foundation for the Advancement of Science is an indication of the importance of this project internationally.

Easy original-script access has been a crucial need of the RLG East Asian, Jewish, and Middle East Studies communities for years. RLG's *Eureka* searching interface can now display the Arabic, Chinese, Cyrillic, Hebrew, Japanese, and Korean scripts included in more than two million records in the RLG union catalog (Smith-Yoshimura, 2000).

In a paper presented at the International Symposium on New Techniques and Applications in Libraries, Xi'an, Chinese People's Republic, 8-11 Sept 88, Karen (1988) pointed out that computer applications in East Asian libraries only became a reality during the early nineteen eighties. Developments of the two major bibliographic utilities, RLIN and OCLC, provided opportunities for libraries to become automated in processing their non-Roman alphabet library materials such as Chinese, Japanese and Korean (CJK).

The following bar chart[*] shows CJK yearly growth in OCLC's WorldCat. The "Chinese," "Japanese," and "Korean" represent the records in WorldCat that may or may not contain CJK characters, but the language element in the fixed field is coded CJK, which is, "chi", "jpn", and "kor", respectively. The "others" are the records that contain CJK characters but the language element in the fixed field is not coded CJK.

The exact numbers for CJK records that contain original scripts are as follows (as of January 2002): Chinese (849,489), Japanese (1,059,969), Korean (157,415), and others (1,539).

[*] Read from bottom to top as follows: Chinese, Japanese, Korean, and Others.

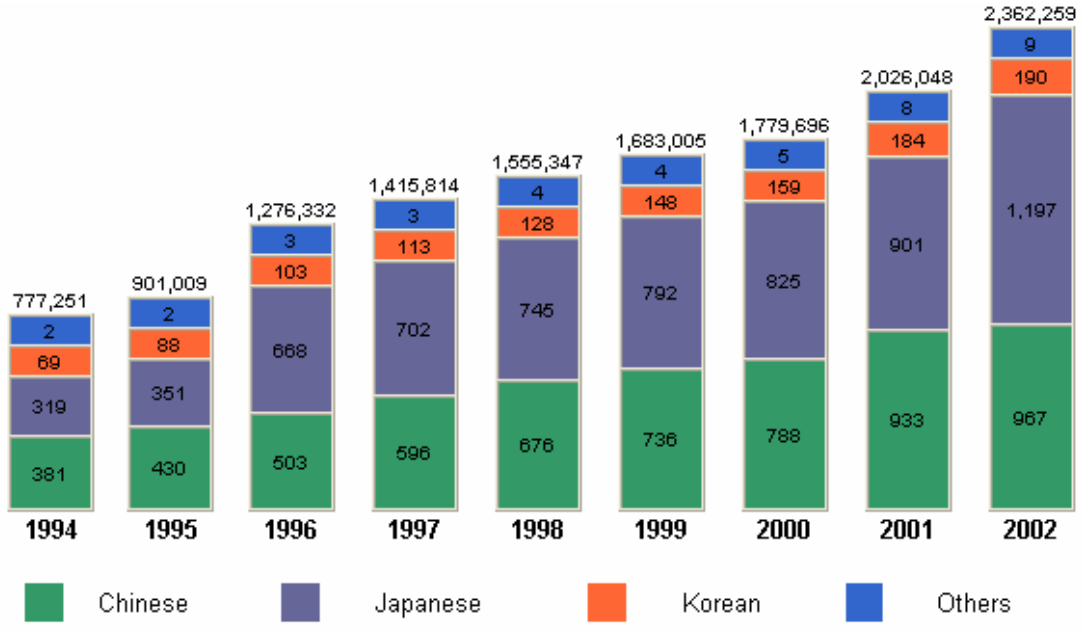

Figure 1.1 CJK Growth in WorldCat (Unit=1,000)
The figure read from bottom to top as Chinese, Japanese, Korean, and Others

(Source: OCLC, CJK growth in WorldCat. Retrieved 226/2002 from http://www.oclc.org/oclc/cjk/stats/cjkyears.htm.)

Tens of thousands of romanized bibliographic records relating to non-Roman script materials exist in the vast OCLC and RLIN databases. Although they started to provide the bibliographic data in vernacular characters, these data are keyed in separately rather than converted automatically from the romanized fields. According to Vassie (1998), this process is time-consuming and therefore costly to produce. However, by producing conversion programs to filter romanized data in various languages back into their original scripts, it is possible to reduce the amount of human intervention to the minimum of quality checking which would normally be required anyway.

Cataloging non-Roman Script Materials in North American Libraries

The cataloging of non-Roman script materials in North America prior to automation was accomplished in a multitude of very divergent ways. Eilts (1996) examined the methods used in

representing non-Roman scripts, which were dictated by local conditions, primarily the staff expertise and equipment available. The challenge now facing the libraries of North America in the area of non-Roman script cataloging, according to Eilts, is the international exchange of data. What is a more logical extension of the desire to maximize the use of already cataloged materials than to begin to acquire this data from countries of origin of the materials? This can be from the book suppliers or the national and research libraries in the various countries.

Elrod (1980), in a paper presented to the Asian Cataloguing Workshop, 20th Biennial Conference, Library Association of Australia, Canberra, Aug., 1979, stated that increasingly, North American libraries in general and Canadian libraries in particular were turning to automated bibliographic databases to support their cataloguing efforts. He discussed in detail the varied and unsettled situation regarding the relation of these libraries' bibliographic records for non-Roman materials to this automation effort.

Converting catalog cards for the online catalog posed many perplexing problems for library collections in Chinese, Japanese, and Korean. One problem is the need for specialized, expensive equipment and complex encoding procedures for inputting and retrieving records with vernacular scripts in those languages. Another is caused by the nature of these languages, whose romanized form is fraught with ambiguity and uncertainty. Given the fairly small size of the Asian collection at San Diego State, for example, the relatively small number of library patrons who have the language ability to use the collection, and the expense of the CJK systems, the decision was made to convert the Asian collections in romanized form only (Rogers, 1986).

Two significant points deserve mention in the San Diego State library experience. The first is, as stated by Rogers, “we had to keep in mind whether or not the system eventually chosen for the online catalog should have a feature allowing it to handle CJK vernacular

characters." The second point is that in order to insure the absolute identification of a title in Chinese, Japanese, or Korean, the library decided that even during the online era, a catalog card for each title in the Asian collection, with information in the original script, would be kept in a file in shelflist or other appropriate order. In cases when library personnel or patrons find the romanized information on the online catalog confusing or misleading, they then can easily refer to this card file containing information in the original vernacular scripts. Until the day when technology has advanced to the point that bibliographical records with CJK characters can be input and retrieved as easily as one can type "abc" on the terminal keyboard without requiring special or expensive equipment, cards with CJK scripts will still be quite essential to accurate bibliographical identification.

The online catalog, the most apparent feature of these automation efforts, has revolutionized the library environment, but for materials in non-Roman scripts the transition to automation has had serious obstacles due to the particular characteristics of these scripts. Until recently the solution for those libraries wishing to automate their non-Roman script holdings has been Romanization, generally based on the Library of Congress Romanization tables. Vernon (1991) examined the advances in computer technology that have provided new possibilities for using these scripts within the online record.

It is generally accepted that the library OPAC means more accessible data for patrons and greater productivity for librarians. While the desirability of automation is not a controversial issue in the United States, some aspects, especially those of non-Roman scripts, remain problematic. However, new developments have opened the door to overcome these obstacles in order to bring these scripts to the mainstream of bibliographic control. These developments are articulated in the following sections.

III. Tools

Character Sets (ASCII, Unicode)

A character set is a list of letters, punctuation, numerals, diacritics, etc. Each character is represented by a number. The American Standard Code for Information Interchange (ASCII) character set, for example, uses the numbers 0 through 127 to represent all English characters as well as special control characters. Unicode is a another character set but, unlike ASCII which uses 8 bits for each character, Unicode uses 16 bits, which means that it can represent more that 65,000 unique characters. This is overkill for English and other Western languages, but it is necessary for some other languages, such as Greek, Chinese, and Japanese. Many computer professionals believe that as the software industry becomes increasingly global, Unicode will eventually supplant ASCII as the standard character encoding.

Fayen (1989) examined inherent difficulties faced by the library community when cataloging non-Roman language materials, especially when it comes to mechanical devices like computers. She examined new problems introduced by library automation because of the lack of a character set that covers non-Roman scripts, and reviewed the ASCII character set, Extended Binary-Coded Decimal Interchange Code (EBCDIC), the ANSEL character set, and the ALA character set developed by the library community to add new codes to the standard ASCII set for characters that are used in other languages. Examples of these characters include the Icelandic thorn (Þ), the Turkish Ĭ, and the alpha (α), beta (β), and gamma (γ) used in scientific texts. In addition, she highlighted the need for terminals and keyboards to support these character sets (e.g., Telex 476L ALA/MARC Keyboard).

Furthermore, Fayen also discussed some problems that may occur by using character sets other than the standard Roman set. Covering three classes of problems; searching problems, sorting problems, and consistency problems, she concluded that "*libraries need to begin to*

define the sort of support needed in a multilingual and multi-script environment so that moving from one to another will be as seamless and transparent as possible for users and for library staff members." And finally, she suggested that "*libraries can help the situation by making sure vendors are aware of their needs and by insisting that new workstations contemplated for purchase have state-of-the-art support for these additional character sets*."

Agenbroad (1991) pointed out that MARC format uses the ALA character set, which was revolutionary when it was introduced because it specified codes for many special characters (e.g., £) and diacritics (e.g., ä, ĕ, and ũ) needed to transcribe accurately bibliographic data in foreign languages. More recently character sets for the Cyrillic, Hebrew, Arabic alphabets and one for Chinese, Japanese, and Korean characters have been added to the MARC format definition. He also provided some options that MARC format can adopt with respect to Unicode:

1. Do nothing. This would be appropriate if vendors do not implement Unicode.
2. Define Unicode as the new MARC character set so every character is 16 bits long.
3. Use an escape sequence to invoke Unicode as needed[*]. This is the technique used in MARC to invoke Cyrillic, Arabic and other character sets.
4. Define fields that would use Unicode exclusively.
5. Dual mode distribution could also be considered. Records could be made available with either the ALA or Unicode character sets at the recipient's option.

In deciding how MARC will respond to Unicode, Agenbroad suggested that we must weigh improved service and reduced dependence on expensive customized devices against the

[*] There are currently two techniques established in MARC 21 to access an alternate graphic character set. One way is a special technique for accessing a small number of characters; the other involves using standard escape sequences to access any well defined character set.

cost of conversion. Other factors include the risks that inaction would further isolate libraries from readers and that a subscriber might convert MARC records to Unicode and market them.

Some views affirmed that libraries with machine-readable records for materials in Arabic, Greek, Hebrew, or CJK should specify Unicode conformity (Why Unicode, 2000), and all libraries should be aware that Unicode is largely supplanting the ASCII character-coding format that has been in use in the United States for several decades. Almost all computer manufacturers, operating system developers, and browser producers have adopted Unicode.

Erickson (1997) showed how the use of Unicode will increase dramatically in the coming few years for many reasons. Use of the Internet has increased awareness of the problems inherent in the use of multiple character set standards. The standard was created by and has been embraced in the computer industry. Many products have come on the market with support for some or the entire standard; Microsoft's use of UTF-8 and Unicode, especially in Windows NT, is a boost to the mainstreaming of such products.

Non-Roman Scripts in Communication Formats (MARC, UNIMARC)

Communication formats are standards for the representation and communication of bibliographic and related information in machine-readable form. These formats are necessary because the information from a catalog card cannot simply be typed into a computer to produce an automated catalog. The computer needs a means of interpreting the information found on a cataloging record. The bibliographic record that is based on a specific communication format, such as MARC, contains a guide to its data, or little "signposts," before each piece of bibliographic information (Understanding MARC Bibliographic, 2000).

The Library of Congress decided to stop filing new entries in its massive card catalogs on January 1, 1980, and to rely primarily on automated data to provide access to the collections.

From that date forward the library possessed two catalogs—the "frozen" card catalog and an online one that included all records in the MARC database and all records cataloged after January 1, 1980, whether available online or temporarily in card format. The automated system, and the underlying MARC format, was not able to handle non-Roman scripts by 1980 (LC, 1977). Therefore, separate card catalogs for these materials were maintained.

According to Lee (1988), the design of the local MARC formats for exchanging bibliographic records and the establishment of the national databases "*seem to have been regarded by librarians in the East Asian countries as top priorities and also as yardsticks for measuring progress in library automation, and all the national libraries in East Asia have made some achievements in this direction*." The first Chinese MARC tape, based on UNIMARC and LC MARC, came out in 1982. All the bibliographic data can be searched in Chinese characters, English, or Romanization. The Japan MARC format is also based on UNIMARC. For processing kanji and kana, it adopts the Japanese Industrial Standard (JIS) character code. Japan MARC tapes were made available in 1981. The National Central Library in Seoul, Korea published the Korean Machine Readable Cataloging (KOR MARC), which has the capability of handling Chinese characters and the Korean hangul, in 1981. KOR MARC also follows the pattern of UNIMARC and LC MARC.

Fung (1983) discussed input and output codes that have been developed for the computerization of Chinese characters. The major codes available include: telegraphic code, three corner coding method, four corner coding method, fixed position method, phonetic symbol method, and root or radical method. Prerequisite formats for the coding of Chinese characters are derived from the revised Chinese cataloging rules, Chinese Character Code for Information

Interchange, Chinese MARC format, and some computerized Chinese character systems of vernacular material have been developed at the Agricultural Scientific Information Centre.

Maruyama (1987) illustrated the efforts of developing the national MARC in Japan, J/Marc, by the National Diet Library (NDL), in order to fit the requirements of Japanese bibliographic information, "*Japanese bibliographic data are usually represented by a combination of Kana (Japanese phonetic script) and Kanji (Chinese characters)*." The improvement of J/Marc, along with standardization of the national character code (i.e., the National Diet Library's NDL-70 Code) helped automation efforts in Japanese libraries to enter a new era of cooperative library networks.

The representation of non-Roman scripts in Roman characters causes information to be distorted in various ways. USMARC now provides for "*alternate graphic representation*," in the 880 'repetitive' field, so that text in the original scripts may be included in bibliographic records (Aliprand, 1992b).

Aliprand (1993) demonstrated that USMARC records that contain non-Roman scripts exhibit two types of linkage between the Roman script fields and their alternate graphic representation (the non-Roman text): linkage based on systematic Romanization, and linkage between names for the same person, place, or thing. The lack of rules for linkage inhibits copy cataloging, and causes inconsistency in record displays. To determine an unequivocal basis for linkage, four types of field association in bibliographic records are examined: hierarchy of components, functional equivalence, semantic equivalence, and systematic Romanization. Semantic equivalence is shown to be the proper basis for linkage. Linkage based on semantic equivalence can be accommodated by the current structure of the USMARC Format for Bibliographic Data.

The success of the MARC format in the United States and in the world led to the development of a number of national and regional formats based on the original MARC format. They are known as MARC-compatible or MARC-like formats, which can accept and regenerate USMARC records without loss of information. Some library automation vendors have also prepared their own formats as extensions to USMARC which consist of additional content designations: fields, subfields, indicator values, and characters for general or local use. OCLC MARC, RLIN MARC, and DMARC (DOBIS MARC) are some examples of extended USMARC formats (Khurshid, 1997). Because of the difference in the MARC communication format (which is based on ASCII) and the DOBIS/LIBIS processing format (which is based on EBCDIC), an interface program is required to translate bibliographic records from ASCII to EBCDIC or from the USMARC format to the DMARC format before the records are loaded into a local DOBIS/LIBIS database. Because of additional content designations, DMARC is considered an extended format of USMARC into which a pure USMARC record would fit, but some of the DMARC format information will not fit into USMARC.

Notwithstanding the availability of USMARC and UNIMARC as international exchange formats, the need for national formats continues to be felt, especially in countries or regions which use non-Roman scripts. Khurshid (1998) reviewed efforts by the Arabian Gulf Region (21 countries) to develop the ARABMARC exchange format for cataloging non-Roman scripts, and provided details of the availability of the Arabic script support in various automated systems in operation in the Gulf Region.

Multilingual Authority Files

Authority files are important tools to ensure that all entries in a database (e.g., library OPACs) have consistent forms of names and subject headings. In a multilingual database, it is necessary

to establish multilingual authority files so that names and subject headings are represented in their original languages and/or scripts and are linked up with other forms, allowing the user to search in his/her language or script.

Raptis and Salaba (1994) discussed the creation and maintenance of three (name, title and subject headings) authority files in the Central Library of Aristotle University of Thessaloniki, Greece. The reasons for establishing bilingual (Greek and English) authority files are explained, and the necessary modifications of the AACR2 rules, which were imperative for the establishment of bilingual headings in the authority files, are represented. The benefits of an OPAC that uses bilingual authority files, which allow one to search in two languages, are described.

As bibliographic files increase in size and complexity, it becomes essential to achieve consistency in forms of name access points for an effective author approach to their contents. In their notes arising from discussions in 1991-1993 within the Standing Committee on Cataloging of the International Federation of Library Associations and Institutions, Division of Bibliographic Control, Murtomaa and Grieg (1994) pointed out that the sharing of identical authorities (established forms of name/title and references) can greatly facilitate international access to entries in catalogs and national bibliographies and helps maximize the benefits of shared cataloging—authority control being possibly the single most costly aspect of cataloging today.

There are many reasons why a single authoritative version of a name cannot always be applied universally. In order to allow exchange in such cases, it would be ideal to have, for each author, at least one form common to existing authority files throughout the world, by which the variant forms could be linked up. Accordingly, the Standing Committee on Cataloging at IFLA

proposes a review of the problems and possibilities for the linking of name authority files internationally, including single-language, and multi-language files.

Landry provided an overview of the Multilingual Subject Access Project that has focused much of its work since May 1997 on establishing equivalences from the LCSH, RAMEAU (Repertoire d ' Autorite-matiere Encyclopedique et Alphabetique Unifie)[*] and SWD/RSWK (The Schlagwortnormdatei/Regeln für den Schagwortkatalog)[**] subject heading languages in two selected subject areas, namely theatre and sports. This work was done in order to test the feasibility of linking headings of three subject heading lists, with the objective of providing the means to access library databases on a multilingual basis. In this context, the project was focused on providing multilingual access to the library users first and foremost rather than catalogers.

As Landry's study focused principally on how the indexing done in one system could be useful to an indexer working in another indexing system, it would appear that there could indeed be some benefits for indexers in having a multilingual access to bibliographic records. To be able to access bibliographical records with the assistance of a multilingual thesaurus would give the indexers a primary insight on the headings used in a different indexing language. This access could be of help in determining the subject content of a document in a language not so familiar to an indexer.

The addition of library catalogs to the mix of information being searched on the Web will open up the Web to focused, topical collections and resources held in and made accessible through the world's libraries.

---

[*] Encyclopedic and Alphabetic Authority Matter ... . The language of indexing used by the National Library de France.

[**] At the Die Detsche Bibliothek, the Schlagwortnormdatei (Subject Authority File, SWD) provides a standardized vocabulary with a controlled terminology. It contains standard and reference forms of subject headings established in accordance with the "Regeln für den Schagwortkatalog, RSWK" (Rules for the Subject Catalogue, RSWK).

Explorations to provide interoperability across multiple authority files, to link and provide switching for displays of authorized headings on an international scale, are underway within the International Federation of Library Associations (IFLA) (Tillett, 2000a). The combinations of Unicode and new technologies are opening up access to all scripts and all languages. Crosswalks, like those provided in OCLC's CORC (Cooperative Online Resource Catalog) project, link Dublin Core (DC) metadata and cataloging rule-based records in MARC and other formats with XML and other communication structures, and expand the opportunities for contributing authority records to an international pool.

Clavel-Merrin (1999) stated that the tasks of creation, management and maintenance of multilingual subject authority files require significant resources, and are therefore best carried out in co-operation. Four national libraries in Europe (the Swiss National Library, la Bibliothèque de France, Die Deutsche Bibliothek, the British Library) have worked together on a feasibility study into linking existing subject authority files in three different languages to offer multilingual subject access to their files.

Tillett (2000b) described some international efforts to provide authority control, including the work of the International Federation of Library Associations and Institutions (IFLA), the AUTHOR Project funded by the European Commission, and related work conducted under the auspices of the International Council on Archives/Committee on Descriptive Standards (ICA/CDS), in Canada. IFLA developed the "Form and Structure of Corporate Headings" guidelines, documented the formulation of names along the lines of national origin in its publication "Names of Persons," and published "Guidelines for Authority and Reference Entries." Attention has shifted from a single authority record for each entity that would be shared internationally through the exchange of records to linking parallel authority records for the same

entity. The access control of the future will account for difference in cataloguing rules, transliteration standards, and cultural differences within the same language as well as for the need for different languages and scripts and will enable users to display the script and form of a heading that they expect.

Project AUTHOR is a shared set of resource national authority files that used selections from the authority files of France, the UK, Spain, Portugal, and Belgium, and tested an adaptation of Z39.50 server software for authority records and displays for user interface. An international standard for authority control records has been developed for corporate bodies, persons, and families. Through joint meetings, efforts have been synchronized to develop authority control at the international level.

Multilingual Interfaces

Most designers and developers have emphasized the concept of user-centered or user-oriented design. One of the most important components of library OPACs, as information retrieval systems designed specifically to be used by end users, is the user interface. The system friendliness is highly dependent on its interface.

Kim and his colleagues examined the correlations between user characteristics and their preferences for selected features of Web-based OPAC systems (Kim, 1999). They concluded that "*system designers should make a more considered appraisal of the users' demographic characteristics in the design of the new generation of OPAC systems such as user-tailored interactive Web-based OPAC systems."*

According to Tyckoson (1991), "*the catalog of the future should be able to meet the needs of searchers who speak languages other than the dominant language of the nation*." In the United States, the population of non-English speaking residents is rising dramatically. Although

many libraries collect materials for specific non-English speaking populations, the cataloguing used for those materials is based in the English language, the so-called "Romanization." Tyckoson further suggested that, "*Help screens and prompts could also be translated from English into other languages*." This type of enhancement will be necessary to meet the demographic characteristics of library users.

Chepesiuk (1997) described the Online Computer Library Center (OCLC) efforts in international projects to increase OCLC online availability, Chinese/Japanese/Korean cataloging systems, Cyrillic alphabet support, and the multilingual interfaces to support all these efforts.

Babu and O'Brien (2000) examined six popular Web OPAC interfaces in use in UK academic libraries (Talis, INNOPAC, WebCat, Voyager, GeoWeb and ALEPH) with an overview of the functions offered via those interfaces. The examination resulted in a checklist indicating the important features and functions that should be offered in Web-based OPACs. Among them are linguistic capabilities: facility to accommodate multilingual libraries, and provision to accommodate non-Roman scripts.

CARL Corporation, Colorado, demonstrated the Spanish and Chinese variations of CARLWeb, its World Wide Web online public access catalog (OPAC) multilingual interface (CARL, 1999).

IV. Rules: Multilingual Issues in AACR2R and ISBDs

AACR2R

The Anglo-American Cataloging Rules-Second Edition, Revised (AACR2R) (Gorman, 1998) includes rules on transcription of non-Roman text in the bibliographic description (for example, Rule 1.0E mandates that the bibliographic description be written in the same script as the source

of information “if practicable”), but AACR2R does not include rules for the formulation of non-Roman access points.

Barry (1983) illustrated the problems encountered when applying AACR2 as interpreted by the Library of Congress to Hebrew and Yiddish personal names. He also discussed some problems relating to the Romanization of such names and proposed some solutions. A plea is made for the eventual development of the capability to handle vernacular records for non-Roman languages in machine-readable form.

In view of the pressure towards standardization in bibliographic description, Terekhova (2000) examined issues relating to entries denoting language, the language of the bibliographic description and transliteration, particularly as they affect Russian catalogers, comparing the Russian Rules with AACR. He also reported the work by OCLC in creating databases accommodating Cyrillic, Japanese and hieroglyphic scripts and the effects of political fragmentation which have increased the number of official languages and described how the Library of Foreign Literature in Moscow deals with collections of materials in about 150 languages.

ISBD

In 1980 IFLA’s Standing Committee of the Section on Cataloguing agreed to review the texts of the ISBDs and it was stated that one of the tasks of the Review Group would be to "*consider the particular problems of non-Roman scripts, and especially those written right-to-left, in order to ensure that all the ISBDs are equally hospitable to publications in all types of scripts*." (Jover, 1987) Consequently, the Review Committee requested and received a number of comments and suggestions from catalogers working with these materials.

In an updated version of a paper presented to the Section on Cataloguing, IFLA Conference, Manila, Aug. 1980, contribution to the ISBD 5-year review, and contribution to a project on problems of applying ISBDs to non-Roman script publications (supported by Unesco contract; final report submitted Feb. 1981), Lee (1981) made suggestions for some modifications to the ISBD format to enable its use for Chinese scripts and outlined procedures adopted in the application of ISBDs in Chinese cataloguing in Hong Kong libraries.

Also in the same convention, Mariam (1981) discussed some of the difficulties of applying ISBDs for monographs, serials, and cartographic materials to the Malaysian national bibliography, which includes publications in a variety of languages and scripts. These include problems related to: prescribed punctuation; parallel titles; and changes in the publication language of serials. She suggested that the IFLA International Office for UBC should give due attention to the problems faced by countries such as Malaysia, that use non-roman scripts from right to left.*

V. Examples of Multilingual Automated Library Systems

This section highlights some examples of libraries' experiences with automated systems handling non-Roman characters. Though the emphasis is directed toward academic libraries in the United States, international libraries' experiences are also considered as signposts to how libraries, in countries where non-Roman scripts are used, utilized the automated systems to better support non-Roman vernacular records.

According to Wong (1992), East Asian collections in the United States have been less affected by the major trends in library automation, largely because local library automation

* Malaysia also uses a left-to-right Roman script for most common usage.

systems have been unable to handle East Asian scripts. This has forced East Asian collections to maintain redundant manual (e.g., card catalogs) and computerized systems, at a great loss of efficiency. Therefore, there is a need to improve local systems to handle vernacular scripts. He concluded, "If the East Asian collections can overcome this obstacle and abandon the traditional catalog cards, their services can enter the mainstream of American academic librarianship."

Khurshid (1992) presented various options considered for developing the Arabized version of DOBIS/LIBIS. Following the implementation of the system, King Fahd University of Petroleum and Minerals considered the following three options for handling Arabic script materials: 1) creating records for Arabic documents in romanized form using the English version of DOBIS. 2) entering the Arabic script data using the English version of the system, or 3) developing an Arabized version of the system for handling bibliographic information in the Arabic script and keeping the Arabic files separate from the English ones. In view of the serious problems with the first two options, the third option was found to be the most appropriate. To prepare the Arabized program, the following steps were taken:

1. Adopted Arab Organization for Standardization & Metrology (ASMO) 449, the standard coding system for the Arabic language to translate Arabic information consisting of letters and numbers to equivalent binary numbers.
2. Wrote some programs in PL/1 and Assembly languages and translated screens, maps, and code tables of the English version of DOBIS/LIBIS.
3. Prepared special files for storing various segments of bibliographic information.
4. Prepared a diacritical table to handle special characters not supported by the IBM X-BASIC terminal.

5. Introduced necessary modifications to LIBIS-batch programs to print catalog cards, spine labels, accessions lists, and notices in Arabic.

Although users were very excited about the catalog because it has almost the same features as the English catalog, there were some limitations related to both the hardware and software (Khurshid 1992). The hardware limitations were mostly related to the keyboard, which didn't fully support the various diacritical characters extensively used in the Arabic script. The software problem was related to the input, sort, and display of the Arabic definite article " ال ". If " ال " is not ignored in sorting, it would result in a large number of entries being clustered together in the file and would impede searching.

Butcher (1993) highlighted the development of the new building of the British Library and its new computer system. Some of the requirements of the online catalog were presented, among which was the ability to deal with over 800 different characters. He discussed managing the special characters in the library's computer system, including entering and displaying these characters. In addition, he talked about the development of help messages that can appear in a number of languages, as well as the future for expanding the system to handle more characters.

A thorough review of the capabilities of automated library systems as they are being used in Saudi Arabia, focusing in more detail on the aspects of processing of bi-lingual information, has been provided by Ashoor and Chaudhry (1994). Using a detailed questionnaire, a survey of currently available library bi-lingual software was conducted. They indicated that even though the survey was limited to Saudi Arabia's institutions, due to practical reasons, the results are expected to be of interest to a wider community of library and information specialists in the entire Arab world as the programs reviewed are used in different countries in this region.

The DOBIS/LIBIS automated library system includes a highly sophisticated module to control automated circulation. Libraries in the Middle East needed to Arabize LIBIS. These facilities were not yet available at King Abdulaziz University, and it was very difficult to predict when they would be. For this reason, a local automated bilingual circulation system (ABCS) using personal computers local area network (LAN) was developed (Iskandarani, 1992). The system was designed to be menu driven, reliable, and simple to use with fully bilingual—Arabic and English—capabilities. The system was also capable of linking with the DOBIS/LIBIS system. The circulation system presented was developed to meet an immediate need of a university library with Arabic and English language collections.

As the only country in the world in which Hebrew is the official language, Israel found it necessary to develop software enabling its research library catalogs from the outset to handle two alphabets—Hebrew (including Yiddish and Ladino) and Roman characters. Starting in 1981, ALEPH, Israel's Research Library Network, utilized locally developed software that could provide both a Hebrew and Roman mode (Lazinger, 1996). However, since the nation's research libraries had large collections in Arabic and Cyrillic languages, an urgent need arose for a system that could also handle Arabic and Cyrillic materials. This led to the development of soft fonts* software instead of the hardware-based Hebrew-English solution that was incorporated in ALEPH's earlier versions. The soft fonts can display on any VT320 or upwardly compatible terminal in Roman alphabet, Hebrew, Arabic, and Cyrillic. Furthermore, in the Hebrew and Arabic modes, the languages of communication with the computer (commands, HELP, and so on) are also in the vernacular.

Working with a foreign vendor and creating a network among three multilingual libraries was a challenging experience for the library of the Jewish Theological Seminary of America

* Soft fonts: a set of characters for a particular typeface that is stored on the computer's hard disk.

(Steinberger, 1994). After four years investigating library systems offering integrated capabilities for English/Hebrew, the library selected the ALEPH system. The intricacies of the Hebrew language posed unexpected problems. Gratz College and the Annenberg Research Institute became key players in the establishment of the network. Several technical difficulties had to be resolved before the interfacing aspects among the three multilingual libraries became functional.

In 1997 Innovative Interfaces, Inc. announced that its INNOPAC Millennium library automation system had the ability to accommodate an additional language that utilizes non-Roman character sets (INNOPAC, 1997). Arabic, added at the request of the American University in Cairo, joined six other languages on the menu: Chinese, Thai, Japanese, Korean, Hebrew, and Cyrillic. The U.S. MARC format specifies that non-Roman characters appear in parallel 880 fields, but INNOPAC Millennium goes beyond that standard. "*The basic design of INNOPAC Millennium allows non-Roman characters to appear in any field, in any record*," according to Steve Silberstein, Innovative's executive vice president.

Developing software for processing bibliographic data of Arabic materials is a relatively recent development. When libraries in the Middle East started automating their collections, most library systems did not provide for the use of Arabic script and this capability had to be developed. Automated library systems started to emerge (like MINISIS, ALEPH, DOBIS/LIBIS, OLIB) to fill the gap for non-Roman scripts. Medawar (1999) described the stages the American University in Beirut (AUB) libraries went through in converting their Arabic materials for use in the OLIB library management system from *Fretwell Downing Informatics* in the United Kingdom. Medawar concluded that although Arabic characters are complex, handling the Arabic materials in an automated library management system has proved to be effective as technical developments made it possible to incorporate them with Roman scripts.

Summary and Conclusions

The attention on non-Roman scripts in library OPACs is likely to increase in parallel to the growing support for creating vernacular bibliographic records in the major research libraries and bibliographic utilities, as well as the increasing trend among automation vendors toward upgrading their systems to become Unicode-compliant. The literature of multilingual access to library catalogs continues to expand rapidly and in many different directions. It is my hope that a unified framework for that literature will emerge. I suspect that such a framework will capture the emerging themes that have been discussed in this review. These are:

- Transliteration as a barrier to information
- Bibliographic utilities' initiatives in creating vernacular bibliographic information
- Supporting vernacular characters in communication formats
- Cataloging non-Roman script materials in North American libraries
- Multilingual authority files
- Multilingual support in character sets
- Multilingual interfaces
- Issues of multilingualism in cataloging rules
- Developments in the library automation industry

It is not too much to hope that, sometime in the near future, the same library OPAC we deal with nowadays will facilitate complete access to vernacular bibliographic records. Future research needs to address the missing pieces. With technologies available and the need emphasized, better coordinated efforts should be directed toward providing the user with a natural way to explore the wealth of information stored in library OPACs.

Empirical Research

This section of the literature review gives an overview of previously conducted research studies on the area of non-Roman scripts in library OPACs. Included are eight studies covering the past three decades, listed below in their chronological order.

Wellisch (1976) reported on a questionnaire survey sent in Spring 1974 to 321 libraries (national, university, public and special) with substantial holdings of works in dissimilar scripts. The libraries represented 55 countries and the response rate was 63%. The results show that Romanization was the virtual equivalent of script conversion. Tables listed the relative ranks of 8 Romanization systems and 'others' for: Cyrillic; Japanese; Devanagari*; Arabic; Chinese; Korean; Greek; and Hebrew. The Library of Congress system was the most widely used for these important scripts, yet its overall rate of application averaged around only 40%. The following were also discussed: Romanized versus separate catalogs by script; Separate catalogs by script and/or language; Catalogs for bi- and multi-lingual communities; and parts of entry which are transcribed and those which are translated.

A survey was carried out by the British Library Lending Division (BLLD) of the transliteration systems for the Russian Cyrillic alphabet in use in 35 national libraries of various European countries together with some other major libraries throughout the world (Smith, 1976). It was found that 13 separate systems were in use, the most heavily used being the *International Organization for Standardization* system and the *German Zentralblatt fur Mathematik* system. It gave an indication of the extensive variations in schemes used and revealed how far from international standardization is the transliteration of only one non-Roman script. However, the

* Also called Nagari, Indian script used to write the Sanskrit, Prakrit, Hindi, and Marathi languages. (Source: "Devanagari" *Encyclopædia Britannica.* Accessed December 13, 2002 from **http://search.eb.com/eb/article?eu=30634**)

increasing use of complex data processing equipment in national library systems throughout the world and the growing demand for the international exchange for bibliographic records make it imperative that recognition be given without delay to the problems of transliteration.

In the early 1970s when the ideas and principles of Universal Bibliographic Control (UBC) gained prominence in the library community, information specialists began to pay more attention to the problem of transliteration. At that time, a survey of UBC acknowledged the "*growing recognition that to promote transliteration or Romanization is at best a limited solution with its own set of problems*" and ascribed the change in attitude to "*the introduction of mechanized systems with non-Roman script capabilities*" (Anderson, 1982).

The RLIN CJK system and the OCLC CJK system are playing an important role in the provision of vernacular bibliographic data for East Asian script materials. Starting with CJK, have they brought the holdings of East Asian collections into the bibliographic mainstream, as they were expected to do? Elman (1991) briefly examined the history and special features of these two CJK systems and reported on a national survey that was conducted among academic and research member libraries in the United States to learn how these systems have been incorporated into their local automated library systems. Findings of the survey indicated that the advantages of the two systems are not fully delivered to the general user because local automated library systems still are incapable of processing this capability in order to integrate fully non-Roman collections into the general collection.

The Library of Congress reconsidered switching the Chinese Romanization system from Wade-Giles to Pinyin in cataloging Chinese language materials in early 1990s. Young (1992) reported a study on library users' familiarity with and preference for the two systems. Five groups of users of Chinese language materials including the "Mainland Chinese," "Taiwan

Chinese," "Singapore Chinese," "Hong Kong Chinese," and "Americans fluent in Chinese" participated in a series of matching tests and responded to a questionnaire. The tests were designed to measure and compare participants' performance in matching Pinyin and Wade-Giles romanized titles with corresponding vernacular Chinese titles. Results indicated the majority scored better in the Pinyin test and preferred Pinyin to Wade-Giles. Research findings supported the Library of Congress proposal for a Pinyin switch.

A wide diversity exists in the current practice of transliterating Cyrillic scripts for use in bibliographic records in online public access catalogs (OPACs). Without knowing which transliteration table was used, it is difficult to retrieve desired records successfully or efficiently. Retrieving an item (e.g., titles or an author's name) from a library's online catalog where it is given only in transliterated form can be a confusing task, even for users who know the Russian language or at least the Cyrillic alphabet. Aissing (1995) explored the problems besetting three groups of Russian language students at Florida University at Gainesville, Department of Germanic and Slavic Languages, faced with romanized Cyrillic bibliographic records, and investigated the students' ability to search the Russian records according to the Library of Congress transliteration table. Results showed the students' success and error rates before and after instruction and established that transliteration is one of the factors limiting access by Russian language students to the Slavic collections.

Vernon (1996) surveyed the automation options available to libraries with Hebrew and Arabic script collections. She also examined the automation decisions that different libraries worldwide have made about automating such collections, particularly considering how their choices relate to overall prioritization and needs assessment at the institution. A library may

choose to Romanize the cataloging data, to use non-Roman script cataloging, or to implement combinations of both. Standards and case studies are provided for each.

The recent decision by the Library of Congress to convert from Wade-Giles to the Pinyin Romanization system was long awaited by many librarians and library users in North America. Arsenault (2002) reconsiders the usefulness of providing romanized-based retrieval in library OPACs for Chinese script materials. During a controlled experiment primarily designed to measure variations in retrieval performance in OPAC title searches based on type of Romanization, thirty library users were asked to use romanized data to search forty monograph titles given to them in Chinese characters. Participants were divided into three treatment groups: WG, monosyllabic pinyin (mPY), and polysyllabic pinyin (pPY). Half of the titles were searched by phrase (exact-title mode) and half were searched by keyword (keyword mode). Transaction logs, capturing the query strings of the participants, were generated by a concealed logging program. The results revealed that while Romanization is an efficient retrieval method that works relatively well for a large number of patrons, it remains problematic for a significant portion of end users who might be better served with character-based retrieval systems.

As discovered from the review of literature above, a survey of current practices and future plans considering the support of non-Roman scripts in library OPACs among academic libraries has not yet been carried out. This leads to the conclusion that a study that deals with this area will fill a gap in the literature and provide guidance to academic and research libraries with regard to non-Roman scripts functional requirements that should be met in their OPACs to better handle the vernacular characters in bibliographic records.

CHAPTER THREE

RESEACH DESIGN AND METHODOLOGY

This study was designed to determine the transition in bibliographic access to non-Roman scripts in library OPACs in selected academic libraries in the United States that are members of the *Association of Research Libraries* (ARL), with specifications for the functional requirements of establishing vernacular characters in library OPACs. To facilitate this study, the following research questions were posed:

3.1 Research Questions

1. What are the characteristics of non-Roman script materials of libraries from the population? How are these materials cataloged, stored, etc.?
2. What is the current status of non-Roman scripts in library OPACs in these libraries? What are their plans for the near future (within five years)?
3. What are the problems facing catalogers and public service librarians with regard to Romanization? And what are the obstacles that hinder vernacular access?
4. What are the functional requirements that could be recommended to designers and developers of library OPACs, as well as librarians, with regard to the support of vernacular characters?

3.2 Delineation of the Problem

The scope of the research problem is further delineated by five ideas regarding non-Roman scripts in library OPACs. These ideas are principle determinants of the focus of the study and to some extent have been factors in the formulation of the research design and methodology.

1. Automated solutions for Roman-script characters are not always applicable to other scripts. Hence library procedures for collections of those scripts may remain unchanged by the introduction of automation.
2. In spite of the increasing need for vernacular characters in library OPACs in terms of the sizeable collections of non-Roman language materials, and with advances in technological solutions for handling non-Roman vernacular characters, bibliographic access to non-Roman characters is still deficient.
3. In view of all issues surrounding Romanization of non-Roman bibliographic data, particularly its effectiveness for retrieval of non-Roman language materials, libraries are not attempting to find better ways to present these scripts in their OPACs. And as bibliographic utilities turn more to supporting cataloging of these scripts in their own vernacular characters, multilingual support in library OPACs will become of increasing value.
4. In academic and research libraries, where electronic resources such as indexing and abstracting services are indispensable, and where a library catalog becomes a part of a networked setting that requires interfacing among all the networked resources (e.g., the linking from a database of journal articles to a journal's record in the library catalog), the library OPAC may not function properly.
5. The problem of handling vernacular characters in library OPACs has not been a top priority in libraries' agendas for decades. This places the burden on the shoulders of OPACs developers and systems librarians to find solutions to multilingual computing problems that will arise upon deciding to tackle this issue.

3.3 Methodology

A qualitative, multi-method study has been undertaken to find out the current developments in the area of non-Roman characters in library OPACs. The multi methods for data collection include employing a national survey of selected ARL academic libraries, contacts via email, and document/Web sites analysis.

A self-administered mail questionnaire has been used as the instrument for investigating current status related to the support of non-Roman scripts in library OPACs in academic and research libraries. Slater (1990) pointed out that the questionnaire is a "*convenient way to obtain data from a large population or sample, particularly if the population is geographically dispersed and travel is not feasible for one reason or another*." According to Powell (1997), there are several advantages of using a questionnaire over other instruments, including "[*Questionnaire] tends to encourage frank answers; helps to eliminate interview bias; and is usually inexpensive to administer*." Therefore, the researcher believed that the survey would be the most appropriate methodology for obtaining the needed research data.

In addition to the survey, as the main methodology in this study, the researcher also employed other techniques for the purpose of collecting data from the other two important stakeholders for this study: the bibliographic utilities, and automated library systems' vendors. The researcher conducted email contacts with these two groups as well as document and Web sites analyses. The goal was to explore their developments with regard to the support of non-Roman scripts in bibliographic records maintained and accessed in library OPACs.

During the course of collecting data for this study, several contacts were established with the technical support staff in the two major bibliographic utilities, OCLC and RLIN, as well as vendors in the library automation industry. In addition, the researcher administered a survey in

Spring 2002 to library automation vendors to determine the support of non-Roman characters in their systems (Shaker, 2002). These data collection efforts have been useful for supporting the data collected from the survey of selected academic libraries.

3.4 Data Collection Instruments

The data came from multiple sources and collection methods and included a self-administered questionnaire to selected ARL academic libraries, contacts via email with technical support staff at the bibliographic utilities and automation vendors, and document/Web site analyses. The questionnaire was generated from discussions and ideas found in the preliminary survey of the literature (See, Appendix "A"). The instrument as used in this study was a written, self-administered mail questionnaire designed by the researcher.

The instrument was divided into five questionnaires. There was one for each of the three language groups: CJK, Middle East (Arabic and Hebrew), and Cyrillic. The other two were for the public services department and the library OPAC system department.

Each of the three language groups' questionnaires consisted of three parts. Part I, *the collection profile*, focused on characteristics of non-Roman script materials and was designed to identify practices in organizing and storing non-Roman script collections. Part II, *the OPAC profile*, intended to assess the availability of non-Roman scripts' support features in library OPACs. The researcher identified two kinds of support features: (1) the ability to change the interface language to user preferences, (2) the ability to search for/display non-Roman scripts in the vernacular. Finally, part III, *the staff profile*, aimed to gather related data about librarians working with non-Roman scripts. Data was gathered mainly from catalogers in order to identify the availability of catalogers with language skills and also to determine whether they experience any problems when handling non-Roman scripts.

The questionnaire was designed with almost all close-ended questions, though some partially close-ended and a few open-ended questions were used. It is commonly believed that close-ended questions are easier to code and analyze.

Since the study was trying to trace the shift in processing non-Roman script materials in the vernacular, a particular focus was placed on identifying how developments related to two equally important stakeholders to this study, the bibliographic utilities and automation vendors, influenced the availability of non-Roman characters in library OPACs. Therefore, the researcher relied on techniques of document/Web site analyses, as well as follow-up questions and contacts via email, in order to illustrate whether the emergence and development of major factors, such as original script cataloging in bibliographic utilities and Unicode-compliant automated library systems, were accompanied by turning to vernacular character representation in library OPACs.

The results of these data collection efforts were integrated with the results of the questionnaire in order to determine whether libraries were able to take advantage of these developments and to identify factors that might hinder libraries in benefiting from these developments. Another goal was to construct implications for these two stakeholders about how libraries plan for the near future to support non-Roman characters in their OPACs.

3.5 Pretesting the Questionnaire

To test the validity and clarity of the survey design and procedures, the questionnaire instrument was examined in two phases. The first was done by consulting a group of experts at the University of Pittsburgh's library system (e.g., senior librarians working in non-Roman script materials). The result of this phase was a decision to divide the original questionnaire into five parts, each to be answered by a particular department within the library (e.g., CJK, Middle East, Cyrillic collections, public services department, and information systems department).

The second phase of the pretest was the pilot study. A tentative, small-scale study of the proposed questionnaires was conducted with selected librarians from the University of Pittsburgh libraries. Therefore, the population of the pilot study was similar to the population of the major study, and the procedures were the same as the one described below for the major study. However, the data collected through the pilot study has not been processed and analyzed in the same manner as for the final study. The pilot group of libraries was asked to complete the questionnaire and to comment on its content and format (see Appendix "B"). The pilot study indicated the need for some changes in the questionnaires, for example, the reconstruction the original questionnaire into five parts and taking off some optional answers that will not be selected by any library, such as obtaining copy cataloging or creating original cataloging in vernacular characters only (questions 5 and 6). These changes were made. The results of the pilot study have not been included in the final report of the main study.

3.6 Data Collection Procedures

Draft versions of the five questionnaires were sent to a pilot group of librarians in the University of Pittsburgh library system which houses one of the nation's sizeable East Asian collections. For the purpose of this study, determining "sizeable" collections of non-Roman script materials was based on at least one of the following three criteria: (1) the collection has a specific identity (e.g., East Asian Collection, Middle East collection, Slavic and East European collection), whether it is separated from the entire collection or integrated within; (2) the collection development policy considers "area studies" a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and (3) the library has librarians whose main responsibilities are to acquire/catalog non-Roman scripts.

These librarians were asked to complete the questionnaire and to comment on its content format (see Appendix "B"). Their suggestions were incorporated into a revisited version of the instrument device. This revised version was submitted to the dissertation committee members, and their suggestions were incorporated into the final version of the questionnaire that is available in Appendix "A".

The research study was also submitted to the *Institutional Review Board* (IRB) in the University of Pittsburgh in order to gain approval of compliance with regulations for the use of human subjects in this study. The IRB approval letter is available in Appendix "D". Once approved, the researcher started the data collection by corresponding with the selected academic libraries for data collection purposes in four ways:

First, an advance-notice letter was mailed to all libraries of the sample. Its purpose was to tell libraries they had been selected for the survey and that they would receive the questionnaire in a few days. Second, about one week later, again to all libraries of the sample—a cover letter with slightly more detail on the survey, a questionnaire, and stamped return envelope were mailed. Third, two weeks after the questionnaire was mailed out, again to all libraries of the sample—a follow-up letter thanking those who had responded and requesting a response from those libraries whose return was not sent. Fourth, and finally, a week after the follow-up letter was mailed out a new cover letter, a replacement questionnaire and stamped return envelope was sent to those who had not yet responded.

All responses were treated anonymously; however, each questionnaire was coded by using a numeric code to enable the researcher to send out follow-up letters to non-respondents along with another copy of the questionnaire.

Documents, e.g., automated systems brochures from vendors, have been analyzed, when available, as well as official documents from bibliographic utilities that are related to their projects in original (vernacular) character cataloging. The researcher has looked mainly for development and progress with regard to providing access to non-Roman scripts in the vernacular, in addition to possibilities and requirements for libraries to take advantage of these developments.

The researcher also contacted, via email, the bibliographic utilities and library automation vendors for additional data gathering. The researcher contacted these two stakeholders, when more information was needed, using prepared guidelines (a set of questions) developed from the literature search, comments from the pilot study, and responses to the questionnaire.

3.7 Sample Frame

Academic libraries in the United States with sizeable non-Roman script collections that are members of the *Association of Research Libraries* (ARL) constituted the population for this study. ARL is a not-for-profit membership organization comprising the leading research libraries in North America. The mission of the Association of Research Libraries is "*to shape and influence forces affecting the future of research libraries in the process of scholarly communication. ARL programs and services promote equitable access to, and effective use of recorded knowledge in support of teaching, research, scholarship, and community services. The Association articulates the concerns of research libraries and their institutions, forges coalitions, influences information policy development, and supports innovation and improvement in research library operations*." (Endorsed by ARL membership, October 1994) There are currently more than 120 members.

The Bylaws of the Association specify that membership shall be by invitation "*to major university libraries whose collections and services are broadly based*" and define such libraries as "*those whose parent institutions broadly emphasize research and graduate instruction at the doctoral level and grant their own degrees, which support large, comprehensive research collections on a permanent basis, and which give evidence of an institutional capacity for and commitment to the advancement and transmittal of knowledge*." (ARL. *Bylaws*. As amended October 1985) The criteria for university library members consist of three parts: the first to ensure a similarity of parent institutional characteristics with the current membership; the second to ensure comparability of size; and the third to ensure diversity and significant contribution to the distributed North American collection of research resources (ARL. *Statement on* …).

A "purposive" sample of Academic libraries from all over the United States was selected after thorough examination of the Web sites of member libraries to determine whether they have sizeable collections of non-Roman script materials. Of 120 plus member libraries, the researcher excluded 28 libraries for being either at Canadian universities, or not university libraries. Of the remaining 94 libraries, the researcher also excluded 24 university libraries in the USA, for not having non-Roman script collections. Please see Appendix "C" for more details.

In addition to the careful examination of the libraries' Web sites, the researcher also matched the list with different associations whose members include academic libraries with non-Roman script materials. That matching done by checking the members' directories available online in each association's Web site. The following associations provided information with regard to academic libraries with sizeable collections of non-Roman script materials in the United States:

- Association of Jewish Libraries (AJL)
- East Coast Consortium of Slavic Library Collections

- Middle East Studies Association of North America (MESA)
- American Association for the Advancement of Slavic Studies (AAASS)
- Middle East Librarians Association (MELA)
- The Council on East Asian Libraries (CEAL)

The researcher chose to base the survey on a "purposive" sample that meets a particular characteristic; that is, academic libraries in the USA that have sizeable collections of non-Roman script materials. McBurney (2001) stated that, "*Purposive samples can almost be considered to constitute a population*." The data collected from these libraries may be more valuable than those that would be obtained in a random sample of all ARL libraries.

3.8 Data Analysis

The items in the questionnaire addressed several issues: characteristics of non-Roman script materials; current status of practices related to the cataloging/retrieval of non-Roman script materials; library catalogs' support for searching/displaying vernacular characters; and other data about librarians with responsibilities related to non-Roman collections.

A coding system was designed to assign a value (e.g., numeric code) to every answer in the questionnaire. Such a system facilitated coding the questionnaires by means of expressing, in terms of numbers, all responses that eventually were collected, including the ones that appear on the questionnaire and those that are added after data collection.

Returned questionnaires were tabulated using the program *Statistical Package for the Social Sciences* (SPSS). The advantages of using a computer are that, in addition to entering the data once and performing any number of counts and comparisons, "[they] *provide enormous flexibility in terms of which questions are analyzed and in what combination*" (Salant and Dillman, 1994). Frequency distributions were obtained for all items.

All the variables that represent the items in the questionnaire are discrete or categorical variables. Univariate analysis in the form of frequency distributions are expressed graphically using frequency and percentage visual displays. Since measures of central tendency (Mean, Mode, and Median) and dispersion (Standard Deviation, and Variance) do not describe discrete or categorical variables, these measures have not been employed to describe the study variables. Of the bivariate analysis techniques commonly used to describe the relationship between two variables, crosstabulation has been used for a deeper analysis of some variables. For example, instead of just learning about whether libraries catalog in Romanization or vernacular (univariate analysis), crosstabulation can be used to show whether libraries that catalog in the vernacular are currently being able to provide users with searching and display functionality in library OPACs.

3.9 Scope and Limitations

This survey was concerned with non-Roman scripts in library OPACs. Non-Roman scripts are those scripts currently supported in OCLC or RLIN bibliographic databases. RLIN supports JACKPHY (CJK since 1983), Arabic (1991), Cyrillic (1985), Persian (1991), Hebrew (1988), and Yiddish (1988)); while OCLC supports CJK (since 1986), and Arabic scripts (since 2001).

The survey was limited to only those academic libraries members of the *Association of Research Libraries* in the United States, with sizable collections of non-Roman materials. Once again, for the purpose of this study, sizeable collections are defined according to at least one of the following three criteria: (1) the collection has a specific identity (e.g., East Asian Collection, Middle East collection), whether it is separated from the entire collection or integrated within; (2) the collection development policy considers "area studies" a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and (3) the library has librarians whose main responsibilities are to acquire/catalog non-Roman scripts.

CHAPTER FOUR

DATA ANALYSES

Collecting data for this study required a self-administered mail questionnaire sent to selected academic library that are members in the Association of Research Libraries, in addition to analysis of Web sites and documents for bibliographic utilities and library automation vendors. The data collection was conducted in September 2002. Library questionnaires (n=69) produced 48 responses by the cutoff date: three of these were from libraries that could not participate in the survey for reasons such as a shortage of staff or because they are implementing a new automated system. The overall response rate was 69%. However, the data analyses that follow were performed on the 45 usable questionnaires (65%) returned by the data collection deadline.

These 45 responses were broadly diverse in their affiliations (public and private) and geographic locations around the country. Twenty nine libraries were affiliated with public universities and 16 libraries were affiliated with private universities, based on the Carnegie Classification of Institutions of Higher Educations (2000). Table 4.1 provides a distribution for all libraries in the sample frame according to Carnegie Classification concerning the affiliations of respondent and non-respondent libraries.

Table 4.1 Distribution of the Sample Frame According to the Carnegie Classification

	Public Universities (%)	**Private Universities (%)**	**Totals (%)**
Respondents	29 (42.0%)	16 (23.0%)	45 (65.0%)
Non-respondents	15 (22.0%)	9 (13.0%)	24 (35.0%)
Totals (%)	44 (64.0%)	25 (36.0%)	69 (100%)

The presentation of the data analyses that follows will focus on the main data collection instrument in the study, which is the library questionnaire. The results of Web site and document analyses of bibliographic utilities and automation vendors will be provided throughout the discussion of the answers to the questionnaire, where appropriate.

This chapter is divided into three parts, based on the division of the questionnaire instrument: (1) Bibliographic access to non-Roman script materials, (2) Public services for non-Roman script materials, and (3) The profile of automated library systems. Part one deals with characteristics of these script materials, the provision of romanized vs. vernacular access to the bibliographic records, problems encountered by catalogers with regard to Romanization, and considerations for the future in providing access to original (vernacular) characters in library OPACs. Part two discusses issues related to the provision of public services to library users interested in these script materials: librarians' and users' access to transliteration tables, the provision of bibliographic instruction, the availability of instructional materials, and problems encountered by public services librarians with regard to access to these script materials. Finally, part three highlights the profiles of the automated library systems used in academic libraries with sizeable non-Roman script materials.

4.1 Bibliographic Access to Non-Roman Script Materials

4.1.1 Non-Roman Script Materials

Non-Roman script materials are of interest to academic and research libraries for purposes of teaching and research in academic programs that study the languages and cultures of countries where these scripts are used. The questionnaire data revealed that among the six collections

studied, East Asian materials ranked first, followed by Cyrillic collections. Middle East collections are last. Table 4.2 shows the frequency distribution of these collections.

Table 4.2 Frequencies / Percentages Distribution of Non-Roman Materials

Response	Chinese (%)	Japanese (%)	Korean (%)	Arabic (%)	Hebrew (%)	Cyrillic (%)
Yes	29 (64.4%)	28 (62.2%)	24 (53.3%)	19 (42.2%)	20 (44.4%)	26 (57.8%)
No	14 (31.1%)	15 (33.3%)	19 (42.2%)	26 (57.8%)	25 (55.6%)	17 (37.8%)
Missing	2 (4.4%)	2 (4.4%)	2 (4.4%)	0	0	2 (4.4%)
Total (%)	45 (100%)	45 (100%)	45 (100%)	45 (100%)	45 (100%)	45 (100%)

These collections are shelved in academic libraries either separated from Roman script materials, or integrated within the overall collection. Figure 4.1 illustrates the relation of non-Roman to Roman collections. East Asian collections tend to be separated from the Roman collections, while the Middle East collections take the opposite direction and tend to be integrated. All Cyrillic materials in this sample are integrated with the Roman collections.

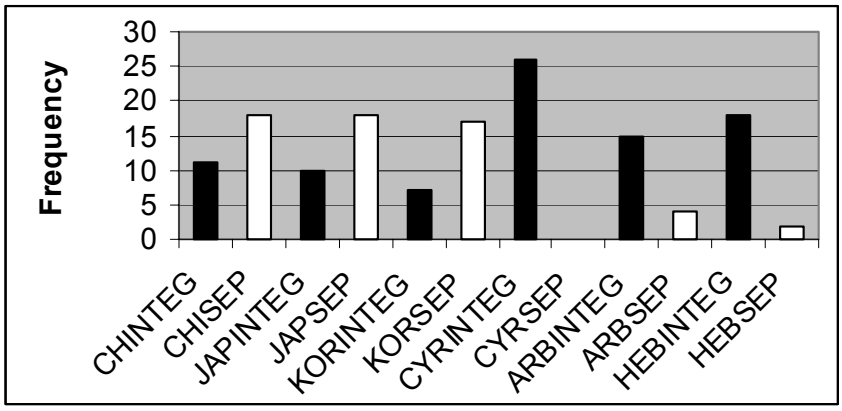

Figure 4.1 The Distribution of Non-Roman and Roman Collections

4.1.2 Source of Bibliographic Records

The three groups of non-Roman script collections did not vary much in their dependences on the two major bibliographic utilities: OCLC and RLIN. Table 4.3 shows the various frequencies for each script collection studied. The East Asian collections rely on OCLC more as often than on RLIN. However, 14-16% of these collections use both of the bibliographic utilities at the same time, an indication that there are still some pluses and minuses of each that mean that some libraries cannot depend entirely on one single utility.

Table 4.3 Frequency Distribution of Sources of Bibliographic Records

	OCLC (%)	RLIN (%)	BOTH (%)	Total (%)
Chinese	18 (62.1%)	7 (24.1%)	4 (13.8%)	29 (100%)
Japanese	17 (60.7%)	7 (25.0%)	4 (14.3%)	28 (100%)
Korean	14 (56.0%)	7 (28.0%)	4 (16.0%)	25 (100%)
Arabic	8 (42.1%)	3 (15.8%)	8 (42.1%)	19 (100%)
Hebrew	10 (50.0%)	3 (15.0%)	7 (35.0%)	20 (100%)
Cyrillic	17 (65.4%)	1 (3.8%)	8 (30.8%)	26 (100%)
Total	84	28	35	

Similarly, the Middle East collections tend to use OCLC more than RLIN. A significant difference here is that a higher proportion of these collections use both of the utilities as source of bibliographic records (35-42%). Finally, the table shows the strong position of OCLC as a source of bibliographic records for Cyrillic script materials, even though RLIN has provided vernacular Cyrillic cataloging since 1985.

4.1.3 Storing Bibliographic Records

The library OPAC in academic libraries became the main gateway for accessing bibliographic records of non-Roman script materials. The percentages of the sample libraries which responded to the survey indicate that fact clearly. Among these libraries, 96.6% have bibliographic records of Chinese materials in library OPACs, 96.4% have records of Japanese materials available in their OPACs, and 95.8% provide access to records of Korean materials in library OPACs. All libraries with Arabic, Hebrew, and/or Cyrillic collections indicate that bibliographic records of these script materials are totally available in the OPACs.

In addition to the library OPAC, a few libraries indicated that some unconverted records are still in a paper file, e.g., Card catalog. However, it seems that those records will end up in the OPAC in the near future, as one library stated: "We still have some cards for older collections and will soon recon them."

4.1.4 Copy vs. Original Cataloging

This section discusses the types of bibliographic records that are obtained (copy cataloging), or are created (original cataloging) for each of the three groups of script materials. The frequency distributions for East Asian collections are presented in the bar chart in Figure 4.2. The majority of respondents perform both copy and original cataloging in the combined romanized and vernacular characters. About 85% of the respondents perform combined copy cataloging. The percentage reaches 90% when it comes to original cataloging in the combination of romanized and vernacular characters.

It should be noted, however, that these high percentages reflect the performance as in the bibliographic utilities, and not in the library OPACs. In other words, the vernacular cataloging is

performed in either OCLC-CJK or RLIN-CJK. As we will see later, the records are loaded in OPACs, but the display of vernacular characters is suppressed. As noted by one library, "We obtain copy in RLIN with vernacular and derive our records in Roman and vernacular but then strip the vernacular when loading into our catalog."

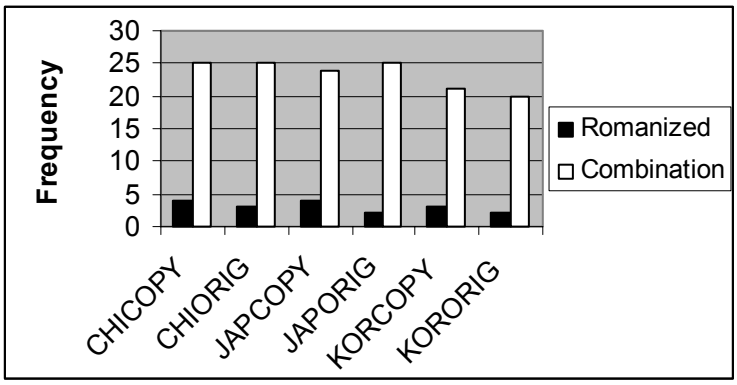

Figure 4.2 CJK Copy vs. Original Cataloging

To better understand whether there is a relationship between the type of CJK copy cataloging and the bibliographic utility where the copy records originate, we need to look at Table 4.4 below. Note that the table is only for Chinese copy cataloging, as a good representative of CJK scripts. The Chinese collection is the most widely available collection among the responding libraries.

Table 4.4 Crosstabulating Chinese Copy Cataloging and Source of Chinese Records

Type of Chinese Copy Cataloging	Source of Chinese Records			TOTAL (%)
	OCLC (%)	RLIN (%)	BOTH (%)	
Romanized	3 (10.3%)	0	1 (3.4%)	4 (13.7%)
Combination of Rom. and Vernacular	15 (51.7%)	7 (24.1%)	3 (10.3%)	25 (86.2%)
TOTAL (%)	18 (62.0%)	7 (24.1%)	4 (13.7%)	29 (100%)

At the far right of the table, notice that 25 libraries or 86.2% of the respondents use the bibliographic utilities to obtain copy records with Chinese characters; only 4 libraries or 13.7% do not. This indicates, therefore, the strong effect of bibliographic utilities as the source for CJK copy records that contain vernacular scripts.

In contrast to East Asian collections, the survey results indicated that the majority of libraries doing Cyrillic cataloging perform it in Romanization, with only two libraries (n=26) obtaining copy records with a combination of romanized and vernacular Cyrillic characters (7.7%). Figure 4.3 illustrates the frequency distributions of these responses.

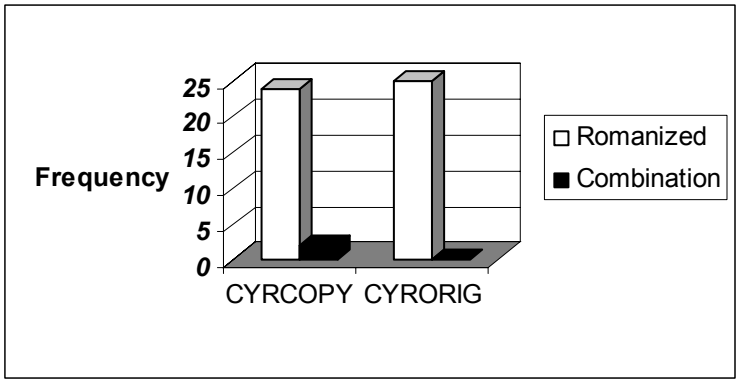

Figure 4.3 Cyrillic Copy vs. Original Cataloging

The crosstabulation on Table 4.5 between Cyrillic copy records and bibliographic utilities shows that 92.3% of the respondents obtain copy records with only romanized Cyrillic data; the other 7.7% get Cyrillic script characters when downloading the bibliographic records from the utilities. This indicates that the two major bibliographic utilities are not utilized to obtain records with vernacular scripts for Cyrillic collections. Among the eight libraries that have access to RLIN, only one library does obtain Cyrillic records with scripts. Since OCLC does not support

Cyrillic vernacular cataloging, we should be suspicious of the answer provided by the library that indicates cataloging of the Cyrillic collection in a combination of romanized and vernacular characters.

Table 4.5 Crosstabulating Cyrillic Copy Cataloging and Source of Cyrillic Records

Type of Cyrillic Copy Cataloging	Source of Cyrillic Records			TOTAL (%)
	OCLC (%)	RLIN (%)	BOTH (%)	
Romanized	16 (61.5%)	1 (3.8%)	7 (26.9%)	24 (92.3%)
Combination of Rom. and Vernacular	1 (3.8%)	0	1 (3.8%)	2 (7.7%)
TOTAL (%)	17 (65.3%)	1 (3.8)	8 (30.7%)	26 (100%)

In the case of Arabic (n=19) script materials, the results indicate similar instances of copy and original cataloging. Among these respondents, 52.6% obtain romanized cataloging versus 47.4% who do a combination of romanized and vernacular Arabic scripts in bibliographic records. The situation is different in Hebrew cataloging. Among the respondents (n=20), 75% perform romanized cataloging, and 25% use a combination of romanized and vernacular Hebrew scripts in bibliographic records. Figure 4.4 illustrates the frequency distributions of these responses.

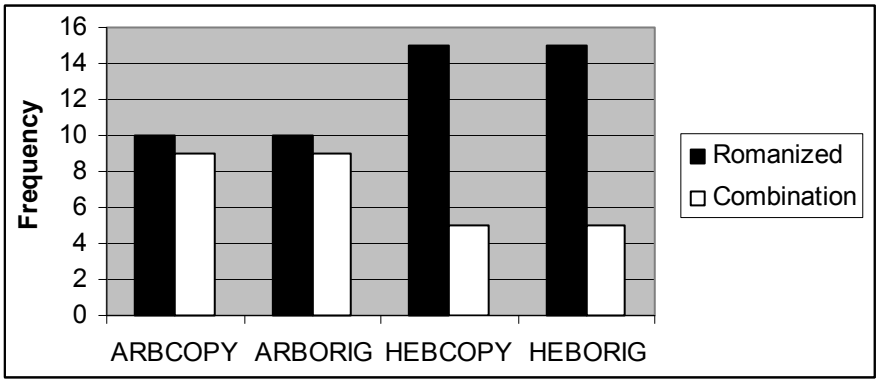

Figure 4.4 Middle East Copy vs. Original Cataloging

The crosstabulation in Table 4.6 below indicates how the bibliographic utilities have been used for Arabic copy cataloging. Arabic script has been chosen because it is supported in both of the utilities; in RLIN since 1991 and in OCLC since 2001.

Table 4.6 Crosstabulating Arabic Copy Cataloging and Source of Arabic Records

Type of Arabic Copy Cataloging	Source of Arabic Records			TOTAL
	OCLC	RLIN	BOTH	
Romanized	5 (26.3%)	0	5 (26.3%)	10 (52.6%)
Combination of Rom. and Vernacular	3 (15.7%)	3 (15.7%)	3 (15.7%)	9 (47.4%)
TOTAL (%)	8 (42.0%)	3 (15.7%)	8 (42.0%)	19 (100%)

On the far right column, 52.6% of the responding libraries obtain romanized Arabic copy records; 47.4% use the bibliographic utilities to get records with Arabic script characters. Note that five of the eight libraries that use OCLC for Arabic cataloging only perform romanized cataloging. This could be explained in light of the recent support for Arabic cataloging in this major bibliographic utility, that started in 2001. It is expected that this number will be reduced in the near future as the service matures.

4.1. 5 Transliteration Schemes

Where Romanization is the only method of cataloging non-Roman script materials, respondents were requested to identify which scheme(s) is(are) being used. There was only one scheme being used for each single script. The reported schemes for romanizing collections are as follows:

- Chinese: PY (Pinyin), changed from Wade-Giles in 2001.
- Japanese: HP (modified Hepburn)

- Korean: McCune – Reischauer
- *ALA/LC Romanization tables* (1991) is the sole scheme that is widely used for Arabic, Hebrew, and Cyrillic scripts.

4.1.6 Reasons for Romanizing Non-Roman Script Materials

Respondents who only catalog non-Roman scripts by Romanization were queried about reasons for doing so (see question # 8). Only seven libraries (n=45) answered that question in the case of East Asian materials. The remaining libraries catalog these collections in both romanized and vernacular characters, and were not required to answer that question. Table 4.7 represents the three main reasons as identified by the respondents. Note that the non-selection of the first reason, "Romanization is adequate for library goals," indicates that those respondents see Romanization as "inadequate" for their library goals. Similarly, these respondents have no problems with obtaining the records with scripts from the bibliographic utilities, meaning that the three remaining reasons constitute the driving force for romanizing East Asian script collections.

Reasons	Yes (%)	No (%)	Total (%)
Romanization is adequate	0	7 (100%)	7 (100%)
No staff with lang. skills	1 (14.3%)	6 (85.7%)	7 (100%)
No support in system	6 (85.7%)	1 (14.3%)	7 (100%)
High cost	3 (42.9%)	4 (57.1%)	7 (100%)
Problems with utilities	0	7 (100%)	7 (100%)

Table 4.7 Reasons for Romanizing CJK Collections

However, other reasons mentioned by the respondents traced the decision to cataloging CJK materials using Romanization back to the pre-automation time, when "Romanization was regarded as the most convenient method for all non-Roman scripts." This indicates that the long-standing tradition with Romanization still has an effect in today's practices.

In the case of the Cyrillic script collections, the reasons respondents (n=26) identified for romanizing the collection varied slightly. While the absence of support to use vernacular characters in automated systems still represents a high percentage (69.2%), a significant proportion of the respondents specified satisfaction with Romanization (73.1%). Table 4.8 delineates the reasons for romanizing that resulted from the survey.

Table 4.8 Reasons for Romanizing Cyrillic Collections

REASONS	YES (%)	NO (%)	TOTAL (%)
Romanization is adequate	19 (73.1%)	7 (26.9%)	26 (100%)
No staff with language skills	4 (15.4%)	22 (84.6%)	26 (100%)
Not system support	18 (69.2%)	8 (30.8%)	26 (100%)
High cost	5 (19.2%)	21 (80.8%)	26 (100%)
Problems with Bib. utilities	10 (38.5%)	16 (61.5%)	26 (100%)

Another reason, as identified by one librarian, was: "Mostly just continuation of pattern of past practice when Cyrillic wasn't an option. Previously, in the manual environment, we did use Cyrillic."

Eighteen libraries that have sizeable Middle East (Arabic/Hebrew) collections answered this question. The reasons specified by those libraries are represented in Figure 4.5. The figure

illustrates that the inadequate support for vernacular characters in the automated library systems was the reason most often selected by the respondents for romanizing Arabic/Hebrew collections (66.7%).

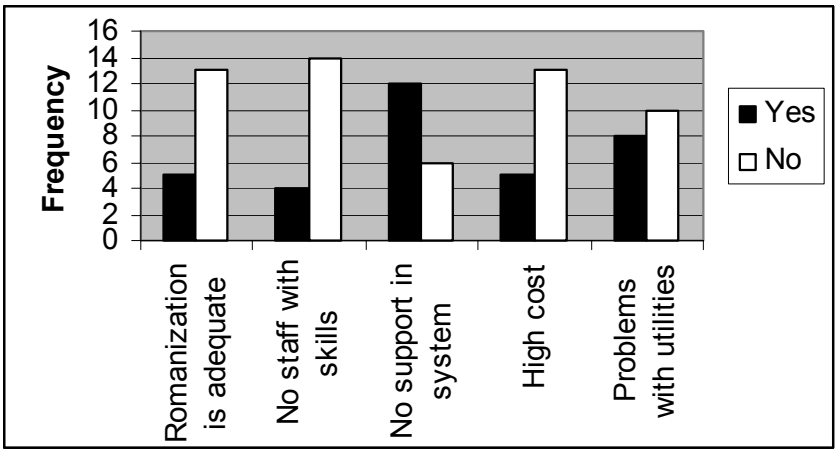

Figure 4.5 Reasons for Romanizing Middle East Collections

Furthermore, it seems that the technological capabilities were a big obstacle that hindered libraries from taking advantage of the innovations in vernacular cataloging. As one library pointed out, "We had hoped to participate in the OCLC Arabic pilot program but that required a Windows 2000 system and we were using Windows 98." The second significant reason (44.4%) was the problems the sample libraries face when obtaining the records from bibliographic utilities. The high cost associated with cataloging in the vernacular was a reason specified by 27.8%. A similar proportion of libraries indicated the adequacy of Romanization for cataloging their Middle East collections.

Finally, 22.2% of the sample libraries catalog these collections in Romanization because of shortage of catalogers with the appropriate language skills. Another important reason stated by one library is the "Lack of staff time to make the transition to Arabic script cataloging in OCLC, and to continue creating dual script records."

4.1.7 Starting Time of Vernacular Cataloging

Question nine of the questionnaire asked libraries that do vernacular cataloging about the starting year of cataloging non-Roman script materials. Only libraries with CJK collections responded to this item. Their input could be divided into two types, summarized below:

1. Libraries that started vernacular cataloging as soon as scripts became available through bibliographic utilities (for CJK, that was mid 1980s), but these libraries do not catalog directly into the OPAC, they catalog in the bibliographic utilities and download the records into the OPAC. Some of these responses are articulated below:
    - We do not catalog directly into the OPAC. We catalog into RLIN and FTP the records into the OPAC.
    - Cataloging in RLIN, not yet in the OPAC.
    - We have not yet started in our OPAC, but we started doing CJK records in RLIN for all CJK languages in November 1983.
2. Libraries that started vernacular cataloging as soon as scripts became available through bibliographic utilities, and started very recently to use their local automated systems. Some of these responses are summarized below:
    - Started vernacular cataloging in OPAC in 2002, but in OCLC in 1985.
    - We started cataloging in RLIN in 1989 with vernacular scripts. These scripts were not displayed in our local system until 2001 when we acquired a new automated system with the capability to display.
    - We were inputting Japanese characters into OCLC long before we could display/search them in the OPAC. We used OCLC for searching/verifying bibliographic information (including the public).

4.1.8 Reasons for Cataloging in Vernacular Scripts

Question 10 of the questionnaire asked all libraries that catalog in the vernacular to identify the reasons that encouraged them to do so. Table 4.9 shows the frequency distributions and percentages for these reasons as identified by respondents to the East Asian part of the questionnaire (n=25).

Table 4.9 Reasons for Vernacular CJK

REASONS	YES (%)	NO (%)	TOTAL (%)
BETTER ACCESS	22 (88.0%)	3 (12.0%)	25 (100%)
LIBRARIAN AVAIL	18 (72.0%)	7 (28.0%)	25 (100%)
SYS. SUPPORT	14 (56.0%)	11 (44.0%)	25 (100%)
RECORDS AVAIL	21 (84.0%)	4 (16.0%)	25 (100%)

Providing better access to users of the collection was the most often selected reason, as stated by one library, "Vernacular data is crucial for identification and retrieval of CJK language materials." However, it should be noted that better access here mostly refers to filling a functional need for the future, since many libraries have automated systems that do not currently allow for searching and/or display of vernacular. Please refer to Section 4.1.11. Following "better access" was taking advantage of the bibliographic utilities as they started to provide access to bibliographic records in the vernacular. Having librarians with the required language skills was also a determining factor (72%) that encouraged those libraries to catalog in the vernacular. Note that the lack of staff was the least specified reason (14.3%) for romanizing CJK collections (see Table 4.5 above).

As the majority of the Cyrillic collections are being cataloged in romanized form (see Figure 4.3 above), there were no answers to this question in the Cyrillic part of the questionnaire.

Only ten libraries with sizeable Middle East script materials provided input with regard to the reasons for providing access to the vernacular characters. Figure 4.6 shows the frequencies for these reasons.

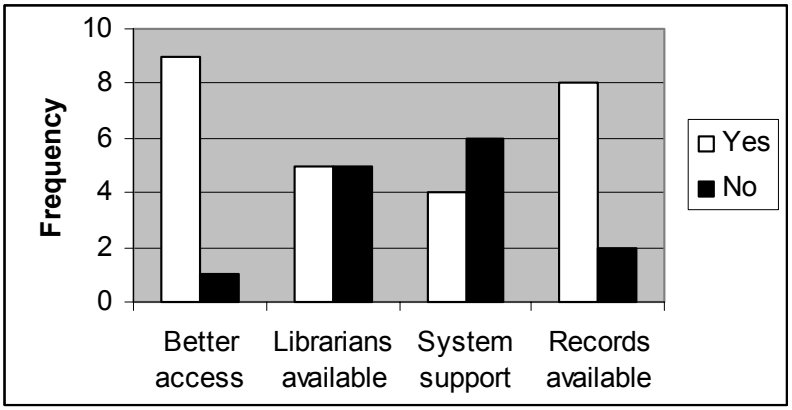

Figure 4.6 Reasons for Vernacular Arabic/Hebrew

As with the CJK collections, providing better access to users of the Middle East collections was the most highly selected reason (90%). The next in order was that bibliographic utilities make available bibliographic records with the vernacular characters (80%). The third and fourth reasons were having librarians with language skills (50%), and implementing automated system with vernacular support (40%).

It should be noted before closing the discussion on these two points (reasons for Romanization and reasons for vernacular) that the lack of system support of vernacular scripts in the automated library system topped the list of reasons for romanizing non-Roman script materials. Meanwhile, it was the least mentioned reason selected by libraries that provide access to bibliographic records in the vernacular. This obviously reflects the idea that technology is a means to an end, which is providing better access to library users.

4.1.9 Interface Languages of Automated Library Systems

Question 11 of the questionnaire asked whether automated library systems permit choosing the interface language based on the users' (librarians and end users) preferences. Figure 4.7 illustrates the frequency distributions for each script of the non-Roman collections studied. Almost none of the automated systems in academic libraries that responded to the survey support the capability to change the interface language to the users' preferences. Note that there are two libraries with a CJK collection, and another library with a Cyrillic collection that indicated that their automated system, Innopac, supports these languages on the interface level.

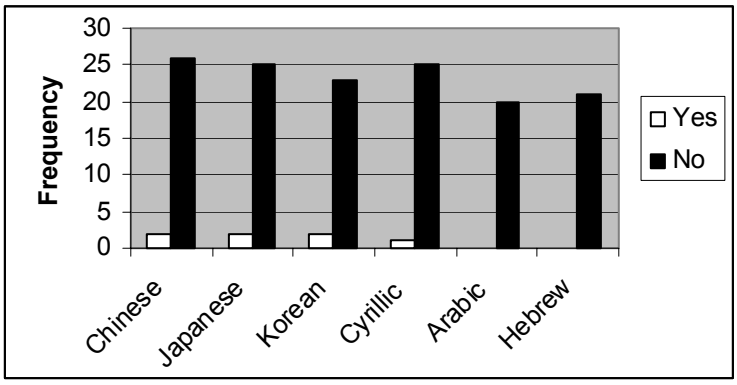

Figure 4.7 Language Interface Support in Automated Systems

Nevertheless, among the "No" responses, one library commented, "The system vendor offers the options of multilingual interfaces, however, our library did not purchase the feature."

4.1.10 The System's Capability of Typing in Vernacular

Respondents were also asked whether their automated library systems, as currently used, permit entering bibliographic data of non-Roman script materials in the vernacular characters (the

cataloging module). Figure 4.8 shows the frequency distributions for each of the vernacular scripts that are of interest to this study.

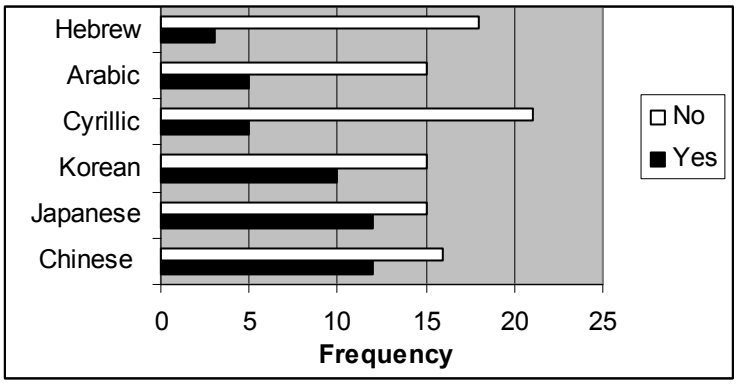

Figure 4.8 Automated Systems Capabilities of Vernacular Typing

The systems' capabilities varied for each of the three groups of vernacular scripts. The support for entering CJK scripts in the cataloging modules was the highest, with 42.9% for Chinese, 44.4% for Japanese, and 40% for Korean. The capability to enter Arabic vernaculars when cataloging Arabic script materials was fourth (25%). About 19% of the respondents determined that their automated systems support the entering of Cyrillic script. And finally, 14.3% of the respondents indicated that they could enter Hebrew characters in the cataloging modules of their automated systems.

Some comments from the respondents are worth noting here: (1) While this capability is not supported in some current systems' implementations, some libraries showed interest in it, as one library stated, "Library implementing new LMS [Library Management System] in 2003 which will allow this." (2) On the contrary, other libraries have not used this function in their systems and have decided to perform their cataloging in the bibliographic utilities. And (3) Some libraries perform this only on the bibliographic utilities' systems as their local systems do not

support entering vernacular scripts: "We build Arabic/Hebrew records in RLIN. [Our integrated library system] does not have vernacular input/edit/display."

Table 4.10 represents the crosstabulation of automated systems' CJK typing capabilities and the type of original cataloging performed in academic libraries that responded to the survey.

Table 4.10 Crosstabulating Automated Systems' Capability of Typing CJK Vernaculars and Type of Original Cataloging

Type of Original Cataloging	Typing CJK Vernaculars in Automated Systems					
	Chinese		Japanese		Korean	
	Yes (%)	No (%)	Yes (%)	No (%)	Yes (%)	No (%)
Romanized	0	3 (11.1%)	0	2 (7.6%)	1 (4.7%)	1 (4.7%)
Combination	12 (44.4%)	12 (44.4%)	12 (46.1%)	12 (46.1%)	7 (33.3%)	12 (57.1%)
TOTAL (%)	12 (44.4%)	15 (55.5%)	12 (46.1%)	14 (41.9%)	8 (38.0%)	13 (61.8%)

All libraries that indicated their systems support typing CJK scripts perform original cataloging with vernacular scripts; with the exception of one library that performs romanized original cataloging for its Korean collection. Another important observation in Table 4.10 above is that the majority of libraries that have systems without CJK typing capabilities also perform original cataloging using scripts. The comments obtained from these libraries clarified that such original cataloging is performed in the bibliographic utilities, and in this case, there are three situations:

1. The records with scripts are downloaded from the utilities to be displayed in the OPAC, if the local system allows displaying,
2. The records with scripts are downloaded from the utilities, but scripts are stripped off because the local system cannot display them, or

3. These libraries provide end user access to special workstations that are connected to the bibliographic utilities.

Table 4.11 represents the results of the crosstabulation of automated systems' Cyrillic typing capabilities and type of original cataloging for Cyrillic collections.

Table 4.11 Crosstabulating Automated Systems' Capability of Typing Cyrillic Vernaculars and Type of Original Cataloging

Type of Original Cataloging	Typing Cyrillic Vernaculars in Automated Systems		Total (%)
	Yes (%)	No (%)	
Romanized	5 (20.0%)	20 (80.0%)	25 (100%)
Combination	0	0	0
Total (%)	5 (20.0%)	20 (80.0%)	25 (100%)

It is noted that none of the libraries that responded to the survey is performing original cataloging with Cyrillic characters, although one fifth of them have automated systems that support typing Cyrillic characters.

Finally, Table 4.12 represents the results of the crosstabulation of automated systems' Arabic/Hebrew typing capabilities and type of original cataloging for these script collections.

Table 4.12 Crosstabulating Automated Systems' Capability of Typing Arabic Vernaculars and Type of Original Cataloging

Type of Original Cataloging	Typing Middle East Vernaculars in Automated Systems			
	Arabic		Hebrew	
	Yes (%)	No (%)	Yes (%)	No (%)
Romanized	3 (15.7%)	7 (36.8%)	2 (10.0%)	13 (65.0%)
Combination	2 (10.5%)	7 (36.8%)	1 (5.0%)	4 (20.0%)
Total (%)	5 (26.2%)	14 (73.6%)	3 (15.0%)	17 (85.0%)

Table 4.12 above shows complicated results: Three libraries that have automated systems with Arabic typing capability do not utilize this capability and perform original cataloging by Romanization only; on the contrary, seven other libraries or 36.8% do not have this capability in their systems but perform original cataloging with Arabic script characters, in the bibliographic utilities. With regard to Hebrew collections, there are very similar figures. Two libraries or 10.0% of those which have the Hebrew typing capability in their systems, catalog only in Romanization, while about one fifth of these libraries or 20% whose automated systems lack this capability perform original cataloging using Hebrew characters.

4.1.11 The System's Capability of Searching/Displaying Vernacular

Another feature explored in the questionnaire (Question 13) was the automated systems' capabilities to allow end users to search in and display bibliographic records of non-Roman script materials in the vernacular. Figure 4.9 illustrates the frequency distributions of these features in East Asian scripts.

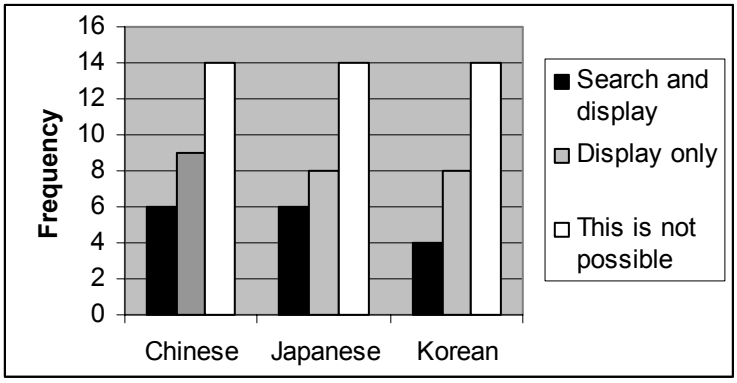

Figure 4.9 Searching and Displaying CJK Scripts

The optimum is, of course, the system's capability to allow both searching and displaying. About 20% of the respondents indicated that they can do that in Chinese (n=29) and Japanese (n=28), while only 15.4% indicated that they have this feature in Korean (n=26). One library stated that, "Searching is possible. We are testing the feasibility; have not implemented." The next best capability in this feature will, then, be the system's support for, at least, the display of vernacular scripts in the bibliographic records. Approximately 30% of the respondents affirmed this capability in their automated systems for the three East Asian scripts. Even so, some libraries encounter problems with this capability, as noted by one librarian: "Character set does not cover all characters. As a result, some characters are unable to display on screen."

Finally, the absence of these features was significant in the automated library systems, where 48.3% of the respondents indicated that their end users cannot search or even display CJK in their OPACs. However, some libraries are working on this, as pointed out by one librarian, "[Search and display] will be possible soon." Another library's solution to overcome the limitations of its automated system is to provide access to their bibliographic records with scripts in their own file in the bibliographic utility. One library noted, "We have RLIN-CJK workstation for the use of readers."

Table 4.13 Crosstabulating Automated Systems' Capability of Searching/Displaying Chinese Vernaculars and Type of Chinese Copy Cataloging

Type of Copy Cataloging	Systems' Capability to Search/Display Chinese			Total (%)
	Search and display	Display only	This is not possible	
Romanization	0	0	4 (13.8%)	4 (13.8%)
Combination	6 (23.1%)	9 (31.0%)	10 (34.5%)	25 (86.2%)
Total (%)	5 (23.1%)	9 (31.0%)	14 (48.3%)	29 (100%)

Table 4.13 above shows the crosstabulation of automated systems' capability to search and display Chinese characters and type of copy cataloging. Of the 29 libraries that obtain copy records with Chinese characters, only six have automated systems that can search and display, nine libraries can display only, and ten libraries do not have systems with these capabilities.

The situation of Cyrillic script (n=26) is significantly different. The majority of respondents indicated the absence of this feature in their OPACs, while the capability to only display vernacular Cyrillic constitutes 12%, and the best deal, searching and displaying, represents about 4%. Figure 4.10 illustrates these percentages.

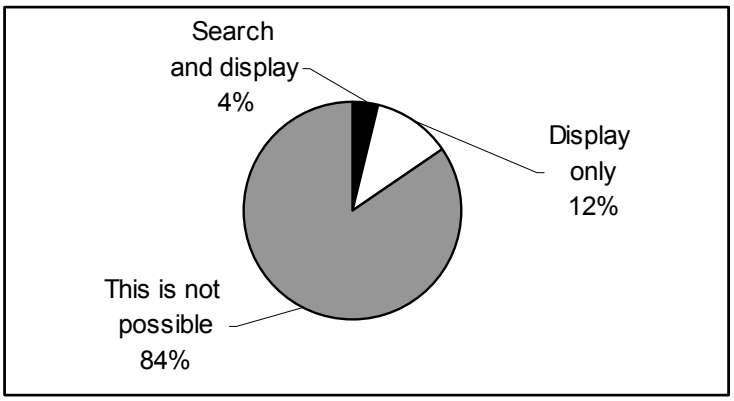

Figure 4.10 Searching and Displaying Cyrillic Vernacular

Two comments from Cyrillic librarians are noteworthy: (1) "The system will display Cyrillic in bibliographic records, but we don't [want to]." On the other hand, (2) "We are testing the indexing and searching capabilities."

Table 4.14 below represents the crosstabulation of automated systems' capability to search/display Cyrillic characters and type of copy cataloging.

Table 4.14 Crosstabulating Automated Systems' Capability of Searching/Displaying Cyrillic Vernaculars and Type of Cyrillic Copy Cataloging

Type of Copy Cataloging	Systems' Capability to Search/Display Cyrillic			Total (%)
	Search and Display	Display only	This is not possible	
Romanization	0	2 (7.7%)	22 (84.6%)	24 (92.3%)
Combination	1 (3.8%)	1 (3.8%)	0	2 (7.7%)
Total (%)	1 (3.8%)	3 (11.5%)	22 (84.6%)	26 (100%)

Among the 26 libraries that are valid for this analysis, 22 libraries indicated that their automated systems lack the capability of searching and displaying Cyrillic scripts. Therefore, they do Romanization. On the other hand, only two libraries have systems that support search and/or display, and they obtain copy records with a combination of romanized and vernacular Cyrillic.

The case of Middle East scripts is somewhat similar to Cyrillic. Among the respondents, 73.7% and 85% indicated that their OPACs do not support the search and display of Arabic and Hebrew, respectively. Searching and displaying Arabic (15.8%) is more common than Hebrew (5%), and both of the two scripts share the 'display only' feature (about 10%) in respondents' OPACs. See Figure 4.11.

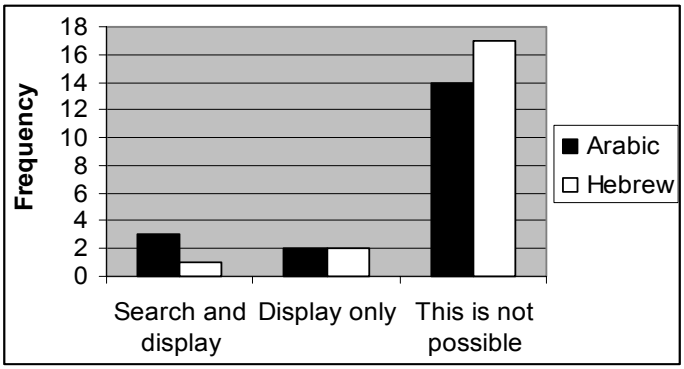

Figure 4.11 Searching and Displaying Arabic/Hebrew Scripts

Table 4.15 is the crosstabulation of the capability of searching and displaying Arabic script and type of copy cataloging. Of the eighteen libraries valid for this analysis, three have automated systems that support Arabic searching and displaying; two obtain romanized copy cataloging, and one does romanized and script copy cataloging. On the other hand, thirteen libraries have systems that lack these capabilities; however, five of them obtain records with scripts.

Table 4.15 Crosstabulating Automated Systems' Capability of Searching/Displaying Arabic Vernaculars and Type of Arabic Copy Cataloging

Type of Copy Cataloging	Systems' Capability to Search/Display Arabic			Total (%)
	Search and Display	Display only	This is not possible	
Romanization	2 (11.1%)	0	8 (44.5%)	10 (55.6%)
Combination	1 (5.6%)	2 (11.1%)	5 (27.7%)	8 (44.4%)
Total (%)	3 (16.7%)	2 (11.1%)	13 (72.2%)	18 (100%)

4.1.12 Non-Roman Characters in Bibliographic Fields

Question 14 of the questionnaire was intended to identify whether the main searchable bibliographic fields are being cataloged in Romanization only, vernacular only, or a combination of romanized and vernacular characters. Although information on subject headings was requested on the questionnaire, the respondents found the instructions unclear with regard to that field and, therefore, gave different answers. The analysis for that particular field will be discussed shortly. Additionally, answers were given only for options 1 and 3 above, hence option 2 (vernacular only) has been excluded in the analysis. Table 4.16 below represents the frequency distributions and percentages for each of the three groups of scripts studied.

Table 4.16 Non-Roman Characters in Bibliographic Fields

Field	Romanization only			Romanized and Vernacular		
	CJK (%)	Cyrillic (%)	Arabic/Hebrew (%)	CJK (%)	Cyrillic (%)	Arabic/Hebrew (%)
Title	6 (23.1%)	24 (100%)	15 (68.2%)	20 (76.9%)	0	7 (31.8%)
Author	6 (23.1%)	24 (100%)	15 (68.2%)	20 (76.9%)	0	7 (31.8%)
Imprint	6 (23.1%)	24 (100%)	15 (68.2%)	20 (76.9%)	0	7 (31.8%)
Series	6 (23.1%)	24 (100%)	15 (68.2%)	20 (76.9%)	0	7 (31.8%)

Of the East Asian group (n=26), many catalog these fields mostly in a combination of romanized and vernacular data (76.9%), while about one fourth of them (23.1%) are doing only Romanization. The Middle East group (n=22) had the opposite position, indicating 68.2% for cataloging these fields in Romanization only vs. 31.8% in combination. Finally, the Cyrillic group (n=24) specified Romanization as the only way of cataloging these bibliographic fields.

With regard to subject heading fields, respondents in the East Asian and Middle East groups, who do catalog in a combination of romanized and vernacular data, indicated that they do catalog the topical subject headings only in English[*], while cataloging other types of subject headings, such as personal/corporate names and geographical places, in both Romanization/vernacular.

4.1.13 Attitudes Toward Vernacular Cataloging

Respondents were asked (Question 15) to indicate their agreement with a statement that read: "Libraries should catalog all non-Roman script materials in the original language and script."

[*] Anglo-American Cataloging Rules have nothing to say about using vernacular scripts in topical subject headings. According to Library of Congress' practice, vernacular data are not allowed in topical subject headings.

Answers were requested only from libraries that catalog their non-Roman collections in Romanization. This question represents a cornerstone in the study because attitudes of catalogers may indicate desired practices other than what is currently being performed for reasons such as shortage of staff or limitations of automated systems. Table 4.17 shows the frequency distributions and percentages for each of the three script groups.

Table 4.17 Attitudes Toward Vernacular Cataloging

Attitude	CJK (n=4)	Cyrillic (n=24)	Middle East (n=15)
Strongly Agree	2 (50.0%)	5 (20.8%)	4 (26.7%)
Agree	2 (50.0%)	9 (37.5%)	9 (60.0%)
Disagree	0	9 (37.5%)	2 (13.3%)
Strongly Disagree	0	1 (4.2%)	0
Totals	4 (100%)	24 (100%)	15 (100%)

Respondents to the CJK questionnaire agreed with the necessity of cataloging all East Asian script materials in their vernacular. The explanation for their attitudes is clustered around the clarity, efficiency, and accuracy that result in ease of access. Their comments are quoted in the list below:

- In CJK languages, there are many synonyms. If only Romanization is present, the users often cannot decipher the meaning.
- I strongly agree but not necessarily ALL CJK. Easier, faster, more accessible, and better serve patrons.
- For the users in the West, it will be convenient to the users if all CJK materials can be cataloged in the original language and script along with Romanization. Frequently, it is cumbersome and difficult to identify CJK bibliographic information based only on Romanization.

- (1) Romanization is very expensive. Users prefer the original language. (2) There are many errors and inconsistencies in Romanization as well as in word division.
- Vernacular data ensures the best bibliographic access, improving identification and retrieval.
- A combination of romanized and vernacular records is the best solution, as it provides better access, better browsing results.
- Cataloging in the original language makes the item most accessible to those who are potential users.

Table 4.18 crosstabulates the attitudes toward CJK vernacular cataloging and same respondents' opinions on "Romanization is adequate to library goals," one of the choices given in question 8 in the questionnaire. All the four respondents on this analysis agree about vernacular CJK cataloging because Romanization is not adequate to their library goals.

Table 4.18 Crosstabulating Attitudes Toward CJK Vernacular Cataloging and "Romanization is adequate to library goals"

Attitudes toward vernacular cataloging	Romanization is adequate		Total (%)
	Yes (%)	No (%)	
Strongly Agree	0	2 (50.0%)	2 (50.0%)
Agree	0	2 (50.0%)	2 (50.0%)
Disagree	0	0	0
Strongly Disagree	0	0	0
Total (%)	0	4 (100%)	4 (100%)

Respondents of the Cyrillic questionnaire were somewhat balanced in their attitudes toward cataloging Cyrillic materials in their vernacular. However, there was a tendency in favor of cataloging in the vernacular (58.3%) against romanized Cyrillic cataloging (41.7%). Some comments from both viewpoints are listed below:

- (Disagree): difficult to find staff in every library who can work with Cyrillic.
- (Disagree): current system configuration presents too many obstacles to doing these. And the product of Romanized cataloging is perfectly adequate.
- (Disagree): Romanization can accurately depict bibliographic information; all patrons and staff are familiar with Romanized data; recataloging is prohibitively expensive.
- (Disagree): expense for hardware, software; user training, long tradition of romanizing.
- Romanization has not proved to be a handicap to the use of collections in Cyrillic script.
- (Disagree): It helps non-speaking/non-reading Cyrillic scripts patrons to identify titles.
- (Strongly disagree): LC transliteration is straightforward enough. Romanization allows access by both library staff and patrons not familiar with the language.
- Bibliographer's view is that transliteration has worked well for decades.
- Cataloger believes it is of some advantage to search and display in original vernacular.
- (Strongly agree): Because transliteration table is another "language" that users have to learn and increase the chances of mistakes and hinders access to materials.
- (Agree): It is the script that is wanted by the users of the materials.
- (Strongly agree): Romanization is not thorough and most romanized catalogs have input and display problems. One Anglo centric alphabet is inaccurate, offensive to many, confusing to others (we use LC transliteration, but many other transliteration systems were used for input in the past).

Table 4.19 is the crosstabulation of attitudes toward Cyrillic cataloging and the same respondents' opinions on "Romanization is adequate to library goals." On the far right column of the table there is an accumulation of 58.3% of Cyrillic librarians who agree with vernacular cataloging because Romanization is inadequate; 41.7% have the opposite position.

Table 4.19 Crosstabulating Attitudes Toward Cyrillic Vernacular Cataloging and "Romanization is adequate to library goals"

Attitudes toward vernacular cataloging	Romanization is adequate		Total (%)
	Yes (%)	No (%)	
Strongly Agree	2 (8.3%)	3 (12.5%)	5 (20.8%)
Agree	5 (20.8%)	4 (16.7%)	9 (37.5%)
Disagree	9 (37.5%)	0	9 (37.5%)
Strongly Disagree	1 (4.2%)	0	1 (4.2%)
Total (%)	17 (70.8%)	7 (29.2%)	24 (100%)

The position of respondents to the Middle East questionnaire (n=15) was very similar to that of East Asian; 86.7% were in favor of cataloging Arabic and Hebrew materials in the respective vernacular scripts while only 13.3% were in favor of Romanization. Some interesting comments worth noting are:

- Arabic and Hebrew Romanization are often confusing to patrons. One needs a strong background and plenty of experience in grammar and Romanization to be able to retrieve relevant records.
- (Agree): There is no one-to-one relationship between Arabic/Hebrew characters and the respective romanized characters. The LC Romanization scheme is non-intuitive. Bibliographic access is already impaired by Romanization.
- (Strongly agree): It would be a good idea to display the language in the script that is designed to do so.

- (Disagree): Scarcity of specialized cataloging staff and lack of interest by students and faculty in vernacular records which can't be supported in existing OPAC.

  The above statement has also the following comment from other library staff:

  > "This section represents the response of our Judaica librarian; many others in the library believe that there is value in having all native scripts available through the catalog."

- (Agree): desirable for library users who know Arabic, etc. to be able to find materials in roman-only setting and Romanization difficulties inherent to languages in Arabic and Hebrew scripts.. no vowels, etc. However, I don't think these materials should be cataloged solely in original script—that requires all staff know all vernacular languages unless separate lib. collections / catalogs are set up, which deviates from principle of union catalogs, and marginalizes these materials.
- (Strongly agree): Most native speakers and professors of Arabic and Hebrew do not use romanized catalogs well, or at all. A combined romanized / vernacular OPAC provides the most accuracy and versatility.

The crosstabulation of attitudes toward Middle East scripts vernacular characters and "Romanization is adequate to library goals" is presented in Table 4.20 below. Ten of the respondents (n=15) or 66.7% agree with vernacular cataloging because they believe that Romanization is not adequate to their library goals. Only two librarians or 13.3% showed the opposite opinion.

Table 4.20 Crosstabulating Attitudes Toward Middle East Vernacular Cataloging and "Romanization is adequate to library goals"

Attitudes toward vernacular cataloging	Romanization is adequate		Total (%)
	Yes (%)	No (%)	
Strongly Agree	1 (6.7%)	3 (20.0%)	5 (26.7%)
Agree	2 (13.3%)	7 (46.7%)	9 (60.0%)
Disagree	2 (13.3%)	0	2 (13.3%)
Strongly Disagree	0	0	0
Total (%)	5 (33.3%)	10 (66.7%)	15 (100%)

4.1.14 Future Consideration of Vernacular Cataloging

In addition to measuring the attitudes toward cataloging non-Roman script materials in the vernacular, the questionnaire was also designed to collect data with regard to academic libraries' future plans in this area (Question 16). Figure 4.12 illustrates the frequency distributions for each of the three script collections.

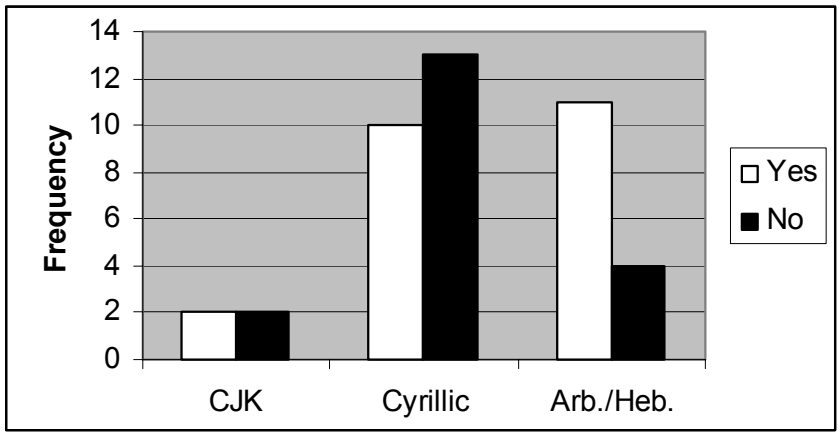

Figure 4.12 Future Consideration of Vernacular Cataloging

Libraries with romanized CJK bibliographic records (n=4) have plans for the near future, within 5 years, for considering vernacular cataloging (50%) vs. 50% for libraries with no future planning to do so. Extensive work is ongoing in some libraries, as indicated by one librarian, "We are working now to figure out proper display of the characters in parallel fields that are indexed and searchable."

Although they catalog their collections only in Romanization (see Figure 4.3), respondents to the Cyrillic questionnaire (n=23) did not indicate significant future plans to consider vernacular cataloging, though it was desired by 58.3% (see Table 4.17). Among the respondents, 56.5% designated no plans vs. 43.5% with plans.

The majority of respondents to the Middle East questionnaire (n=15) indicated that they have future plans for cataloging Arabic and Hebrew script materials in their vernacular (73.3%). Only 26.7% indicated no such plans for the near future.

4.1.15 Obstacles to Using Vernacular Scripts in OPACs

Question 17 of the questionnaire went further to explore reasons why the respondents do not want to catalog their non-Roman script materials in the vernacular characters. Table 4.21 shows the frequencies and percentages for each of the three script groups.

Table 4.21 Obstacles to Using Vernacular Scripts in OPACs

Reasons	CJK (n=4)		Cyrillic (n=25)		Arabic/Hebrew (n=15)	
	Yes	No*	Yes	No*	Yes	No*
No barrier	0	4 (100%)	7 (28%)	18 (72%)	1 (6.7%)	14 (93.3%)
Staff lacks skills	2 (50.0%)	2 (50.0%)	3 (12%)	22 (88%)	1 (6.7%)	14 (93.3%)
Cost	2 (50.0%)	2 (50.0%)	7 (28%)	18 (72%)	5 (33.3%)	10 (66.7%)
No system support	3 (75.0%)	1 (25.0%)	19 (76%)	6 (24%)	10 (66.7%)	5 (33.3%)
Problems w/ utilities	0	4 (100%)	9 (36%)	16 (64%)	4 (26.7%)	11 (73.3%)

* "No" means the choice answer was not selected

The first choice provided in the list of possible answers read: No barrier, we are just fine with Romanization. The three groups indicated "they are not fine with Romanization," with 100% for CJK, 93.3% for Arabic/Hebrew, and 72% for Cyrillic. But note also that 28% of the Cyrillic librarians showed the opposite position. The overall indication, of course, is that Romanization is considered a barrier from the respondents' viewpoints, and, since they cannot do vernacular cataloging, they are forced to live with Romanization.

The second choice provided was: The current staff lacks the language skills. The three groups gave a great indication that this is not the case, with 93.3% in Arabic/Hebrew, 88% in Cyrillic, and 50% in CJK. Note also that half of CJK librarians and 12% of Cyrillic librarians selected that reason. This indicates that the lack of language skills is not the real obstacle to using original (vernacular) characters in library OPACs.

The third choice provided in the list of possible answers read: It may be costly. Although "cost" is not considered an obstacle by 72% of Cyrillic librarians, 66.7% of Arabic/Hebrew librarians, and 50% of CJK librarians, the other views that considered it as an obstacle should not be ignored: 50% of CJK librarians, 33.3% of Arabic/Hebrew librarians, and 28% of Cyrillic librarians.

The fourth choice provided was that: The current OPAC system does not support vernacular characters. Unlike the other choices discussed above, most respondents selected this reason as a major obstacle against using vernacular scripts in library OPACs, with 76% of Cyrillic librarians, 75% of CJK librarians, and 66.7% of Arabic/Hebrew librarians. However, there are about one fourth of Cyrillic librarians, for example, who did not consider that as an obstacle.

The fifth, and last, choice provided in the list of possible answers read: Difficulties with obtaining the records with vernacular characters from the bibliographic utilities (e.g., OCLC, and RLIN). CJK librarians have shown no difficulties at all with bibliographic utilities. This may be interpreted in light of the long history of CJK cataloging in the bibliographic utilities, dating back to the mid 1980s. The other two groups have revealed similar figures with regard to problems with bibliographic utilities. About two thirds of Cyrillic librarians and 73.3% of Arabic/Hebrew librarians indicated they do not experience difficulties when obtaining records with vernacular characters from the utilities. Nevertheless, we still find about one third of Cyrillic librarians and 26.7% Arabic/Hebrew librarians who have such difficulties with utilities.

Based on the above analysis, we can list the possible obstacles in the following order: (1) No system support, (2) It may be costly, (3) Difficulties with bibliographic utilities, (4) Staff lacks language skills, and (5) No barrier, we are fine with Romanization.

Before we close the discussion on this point, some comments are in order:

1. Some libraries did indicate that they have already implemented an automated system that could handle vernacular scripts, but they still experience some problems. As one CJK librarian noted, "[Our automated system] theoretically supports vernacular but we are in early stages of implementation and have imperfect display, no indexing, and no input capability." A Middle East librarian confirmed the same problem, "At the present time we are having difficulty getting vernacular characters to display properly in the OPAC when the records are brought over from the utility." Furthermore, the problems are not only technical, but also managerial. In other words, librarians may need to have a decision from the top management, as declared by one librarian, "No policy decision yet that this should be pursued as we become able to do so with [the automated system]."

2. Other libraries think of bibliographic access to vernacular scripts as a daunting task that might make such endeavor a low priority. As stated by a Cyrillic librarian, "Other priorities may take precedence. Also, retrospective handling of the existing romanized records would be a problem." Similar concerns were echoed by a Middle Eastern librarian, "[It is] time-consuming, especially at beginning. Also, each bibliographic record with dual vernacular/roman info takes more time."

4.1.16 Vernacular Scripts' Access Devices

Respondents who catalog non-Roman script materials in vernacular were asked to give details about the current settings of OPAC workstations that could be used by end users to access the vernacular characters in their libraries (question 18). Figure 4.13 illustrates the frequency distribution for two script groups (CJK, n=16 and Arabic/Hebrew, n=4). There were no instances for Cyrillic in this question, due to the "all romanized" cataloging of Cyrillic collections.

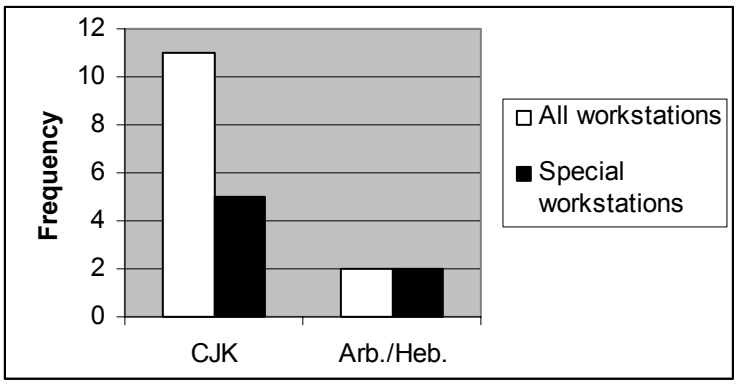

Figure 4.13 Vernacular Characters in Access Devices

The bar chart above shows that about two thirds (68.8%) of libraries with CJK collections provide access to CJK records in all workstations throughout the library. End users can access the vernacular CJK scripts in any OPAC device. The remaining respondents (31.3%) have special devices allotted to provide access to these scripts in particular.

Libraries that provide access to CJK scripts in all OPAC workstations have the following settings in these workstations:

Setting # 1: Using the language capabilities (IME* editor) of Windows 2000, and configuring Windows 2000's "Regional Options" setting for Chinese, Japanese, and Korean, and using Internet Explorer.

Setting # 2: Instructions on screen to switch to CJK port (module), use IE, font Unicode UTF-8, and NJStar CJK Communicator.

Setting # 3: Using separate Web OPAC Port for searching and displaying CJK records, directly linked from Web OPAC opening pages. The original scripts are fully supported by Unicode, but keyword searching is not capable. Special software, such as ROMAX (Conversion applet for Chinese Romanization Pinyin, Wade-Giles, Mandarin Phonetic Symbols II) and WinMASS CJK (a multilanguage software for Windows).

Setting # 4: The record has a toggle [button] that allows display of vernacular characters instead of romanized.

Libraries that provide access to CJK scripts in special OPAC workstations have the following settings in these workstations:

Setting # 1: Special software are installed into special computers: (1) X WINDOW Server** (Hummingbird) for the PC Wins, and (2) CJK xterm, a locally developed software which can display the EAAC [East Asian Character Code] data from Gladis / MELVYL Servers.

Setting # 2: Using a Unicode-compliant automated library system.

The situation is different in Arabic/Hebrew access devices; one half (50%) of these libraries provide access to Arabic/Hebrew vernacular records in special workstations and the other half (50%) provides that access in all OPAC workstations in the library.

Libraries that provide access to Arabic/Hebrew scripts in special OPAC workstations have reported the following settings in their libraries:

Setting # 1: Using Unicode-compliant operating system, such as Windows 2000.

Setting # 2: Providing access to terminals connected to the bibliographic utility.

* Microsoft Office Multilingual Pack **I**nput **M**ethod **E**ditor (IME).
** X Window servers enable transparent access to different operating systems and their applications simultaneously.

4.1.17 Availability of Vernacular Records

Question 21 of the questionnaire asked about the level of availability for bibliographic records with vernacular scripts in the library OPAC. Two groups, CJK and Arabic/Hebrew, responded to that question, while Cyrillic was considered not applicable for this analysis. Figure 4.14 illustrates the frequency distributions for these responses.

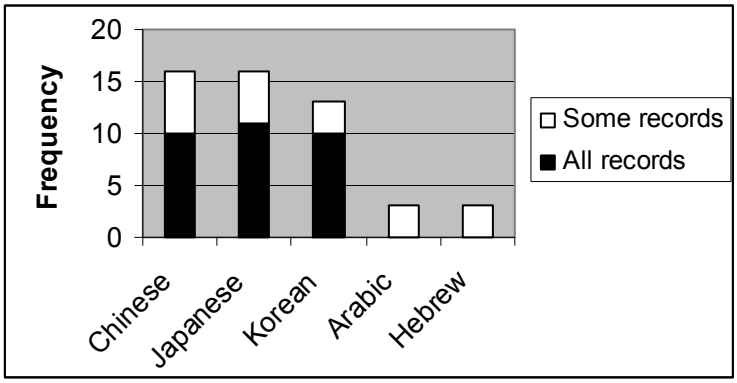

Figure 4.14 Level of Availability for Vernacular Records

CJK scripts are widely available in all bibliographic records, with 62.5% for Chinese (n=16), 68.8% for Japanese (n=16), and 76.9% for Korean (n=13). All Arabic/Hebrew (n=3, each) scripts are available only in some bibliographic records. The remaining bibliographic records are either romanized only, or not converted yet to the OPAC.

Respondents who indicated that CJK and Arabic/Hebrew scripts are available in some of the bibliographic records in the OPAC were requested to identify whether their library plans to convert the other records that do not have scripts in the near future. Table 4.22 lists the frequencies for their responses.

Table 4.22 Future Conversion to Vernacular Records

Scripts	Some Records (n)	Yes	No*
Chinese	6	5	1
Japanese	5	4	1
Korean	3	2	1
Arabic	3	2	1
Hebrew	3	2	1

* "No" means no response

Most of the non-script bibliographic records will be supported by scripts in the near future. However, funds availability was an issue in most libraries. As noted by one librarian, "If funding is available, we will do this."

4.1.18 Non-Roman Script Catalogers

The availability of professional librarians with language skills to deal with non-Roman script materials in academic/research libraries is an important factor when discussing bibliographic access to these collections. Therefore, in question 23 respondents were asked to identify whether they have these types of librarians. Table 4.23 represents the frequency distributions and percentages for their answers.

Table 4.23 Non-Roman Script Catalogers

Librarians with language skills	Yes (%)	No (%)	Totals (%)
Chinese	28 (96.6%)	1 (3.4%)	29 (100%)
Japanese	26 (89.7%)	3 (10.3%)	29 (100%)
Korean	17 (68.0%)	8 (32.0%)	25 (100%)
Cyrillic	22 (84.6%)	4 (15.4%)	26 (100%)
Arabic	17 (85.0%)	3 (15.0%)	20 (100%)
Hebrew	15 (71.4%)	6 (28.6%)	21 (100%)

Academic libraries who responded to this questionnaire seem to be well staffed with CJK, Cyrillic, and Arabic/Hebrew librarians, but their is a shortage of Korean librarians, as indicated by 32%, and also of Hebrew librarians, as indicated by 28.6% of respondents.

But does the availability of catalogers with language skills have an effect on cataloging in the vernacular? The answer is articulated on the bivariate analyses, or crosstabulations, below.

Table 4.24 represents the results for Chinese copy/original cataloging and availability of Chinese catalogers. The strong impact is demonstrated on the 82.7% of the sample libraries that have Chinese catalogers and obtain copy records with vernacular characters. The impact is also demonstrated on the 89.3% of these libraries that create original cataloging with Chinese script.

Table 4.24 Crosstabulating Type of Chinese Cataloging and Availability of Chinese Catalogers

Type of Cataloging	Availability of Catalogers--Chinese Copy Cataloging		Availability of Catalogers--Chinese Original Cataloging	
	Yes (%)	No (%)	Yes (%)	No (%)
Romanized	4 (13.8%)	0	3 (10.7%)	0
Combination	24 (82.7%)	1 (3.5%)	25 (89.3%)	0
Total (%)	28 (96.5%)	1 (3.5%)	28 (100%)	0

The crosstabulation of type of Cyrillic cataloging and availability of Cyrillic catalogers is demonstrated in Table 4.25, which represents completely different positions.

Table 4.25 Crosstabulating Type of Cyrillic Cataloging and Availability of Cyrillic Catalogers

Type of Cataloging	Availability of Catalogers--Cyrillic Copy Cataloging		Availability of Catalogers--Cyrillic Original Cataloging	
	Yes (%)	No (%)	Yes (%)	No (%)
Romanized	20 (76.9%)	4 (15.4%)	21 (84.0%)	4 (16.0%)
Combination	2 (7.7%)	0	0	0
Total (%)	22 (84.6%)	4 (15.4%)	21 (84.0%)	4 (16.0%)

Among 24 libraries that obtain copy records in Romanization, twenty libraries or 76.9% have Cyrillic catalogers with language skills. Furthermore, all 21 libraries whose original cataloging is romanized have catalogers with the appropriate language skills.

Table 4.26 crosstabulates type of Arabic cataloging and availability of Arabic catalogers. The availability of catalogers with Arabic language skills has not impacted the type of cataloging performed. Among the seventeen libraries that are valid for this analysis, eight or 42.1% perform their Arabic cataloging by Romanization whereas nine or 47.4% catalog Arabic materials by both romanized and vernacular characters.

Table 4.26 Crosstabulating Type of Arabic Cataloging and Availability of Arabic Catalogers

Type of Cataloging	Availability of Catalogers--Arabic Copy Cataloging		Availability of Catalogers--Arabic Original Cataloging	
	Yes (%)	No (%)	Yes (%)	No (%)
Romanized	8 (42.1%)	2 (10.5%)	8 (42.1%)	2 (10.5%)
Combination	9 (47.4%)	0	9 (47.4%)	0
Total (%)	17 (89.5%)	2 (10.5%)	17 (89.5%)	2 (10.5%)

4.1.19 Cataloging Problems of Non-Roman Script Materials

The final question in the questionnaire to each of the three script groups requested input with regard to the problems encountered by library catalogers with Romanization of non-Roman script collections. Table 4.27 highlights the frequency distributions and percentages of the responses from catalogers.

Table 4.27 Catalogers' Problems with Romanization

Problems	CJK (n=24)		Cyrillic (n=22)		Arabic/Hebrew (n=19)	
	Yes (%)	No (%)	Yes (%)	No (%)	Yes (%)	No (%)
Multiplicity of schemes	5 (20.8%)	19 (79.2%)	3 (13.6%)	19 (86.4%)	4 (21.1%)	15 (78.9%)
Difficulty w/ scheme	8 (33.3%)	16 (66.7%)	5 (22.7%)	17 (77.3%)	11 (57.9%)	8 (42.1%)
Long time	8 (33.3%)	16 (66.7%)	3 (13.6%)	19 (86.4%)	10 (52.6%)	9 (47.4%)
No problems	9 (37.5%)	15 (62.5%)	12 (54.5%)	10 (45.5%)	5 (26.3%)	14 (73.7%)

Catalogers of the East Asian collections surveyed experienced some difficulties with the three problems listed on the questionnaire. 33.3% of these librarians had difficulties in applying/using the Romanization scheme(s). Some of these difficulties are articulated below. The long time it takes to catalog a book was a special problem encountered by 33.3% of CJK catalogers. About 20.8% of them also expressed the problem of multiplicity of Romanization schemes. But one cataloger noted that, "but now all libraries use Pinyin, so less of an issue." Figure 4.15 illustrates the frequencies of these problems among CJK catalogers.

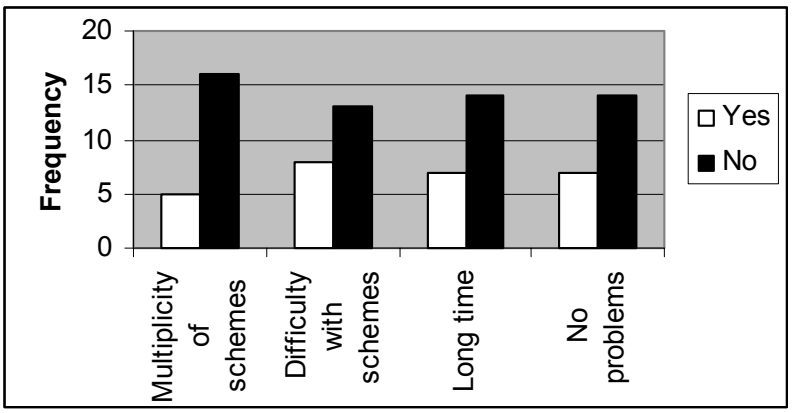

Figure 4.15 CJK Cataloging Problems

Other problems encountered by CJK catalogers, as well as more details on the difficulties of applying/using the schemes, are summarized below:

- Inconsistent application of word division.
- Lack of enough staff with language skills.
- Inconsistent application of Romanization rules in the catalog, which causes difficulty in cataloging as well as on searching. It results in multiple records in OCLC as well.
- Having to change from Wade-Giles to Pinyin has caused a lot of troubles for catalogers.
- The difficulty with Chinese Romanization in RLIN is in applying the Chinese aggregation guidelines to link together syllables forming a word or semantic unit, but this is indispensable for keyword searching in RLIN, when there is no adjacency searching. And the same problem applies to the characters, since we have to space between groups of characters for keyword character searching. Eliminating Romanization would not eliminate the problem.
- Main problem with Japanese Romanization is finding readings for Japanese names, especially for pre-1945 authors.
- The problem with Korean Romanization, for a cataloger who does not know Korean is to find a correct form among various combinations of Hangul Romanization because the pronunciation of consonants depends on their location on the word.

Cyrillic catalogers were the least likely among the three groups to experience any of the three problems specified as optional answers. Among these catalogers, 54.5% indicated no problems with Romanization at all. This could be explained in two ways: (1) the "all romanized" cataloging performed by these catalogers, and (2) the use of just one Romanization scheme, ALA/LC. Figure 4.16 illustrates the frequencies of these problems.

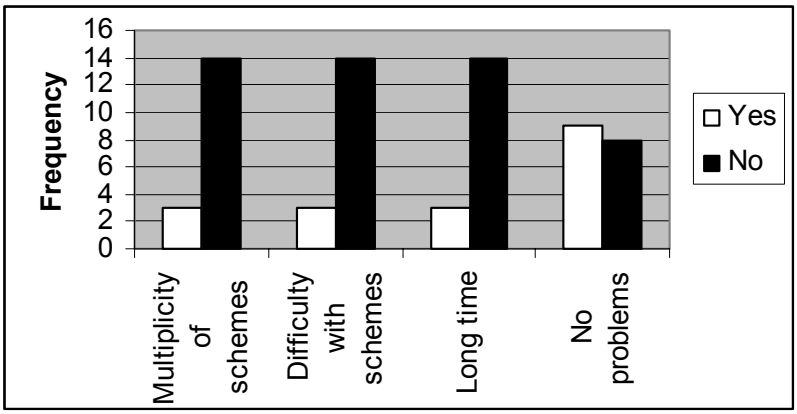

Figure 4.16 Cyrillic Cataloging Problems

In addition to the three problems illustrated in Figure 4.16 above, Cyrillic catalogers did specify some other problems, and these are articulated below:

- Different schemes are used for non-Russian Cyrillic (e.g., Belarusian, Macedonian, and Serbian). Staff are used to only Russian.
- Lack of fluency in transliteration, knowing a language is not the same as being flexible enough to become fluent enough to hop back and forth between Cyrillic and its transliterated forms.
- Staff most familiar with Russian; other languages (e.g., Bulgarian, Ukrainian take longer, more consultation of Romanization scheme.)

Middle East script catalogers were the most likely among the three groups to report problems with Romanization (73.7%). Among these problems were the difficulties in applying/using the scheme(s) (57.9%), the long time it takes to catalog a book (52.6%), and the multiplicity of schemes (21.1%). Figure 4.17 illustrates the frequencies of these problems among Arabic/Hebrew catalogers surveyed.

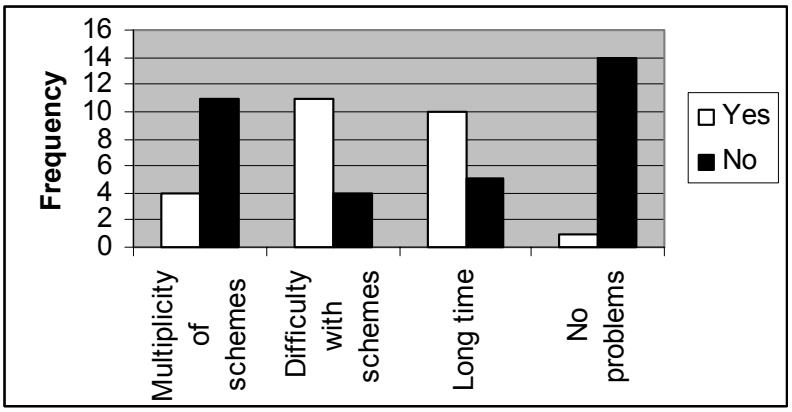

Figure 4.17 Arabic/Hebrew Cataloging Problems

Some catalogers also highlighted the problem caused by Arabic and Hebrew scripts' use of vowels. As commented by one librarian, "Arabic and Hebrew are written without vowels so care must be taken with meaning of all words, and regional variations, especially with names."

4.1.20 Non-Roman Scripts in Bibliographic Utilities

Throughout the questionnaire, a number of questions related to bibliographic utilities. For this reason, and in order to shed more light on these utilities as one of the two driving forces toward vernacular scripts in bibliographic records, this section is devoted to comparing some features and services in the two major bibliographic utilities, OCLC and RLIN.

Table 4.28 below compares the two utilities. It is actually an update of a similar table that appeared in Zeng's dissertation in 1992. Note that the current table covers CJK, Cyrillic, Arabic, and Hebrew scripts, compared to Zeng's focus on CJK only. Information in Table 4.27 was obtained by email communications with the two bibliographic utilities' technical support staff and is augmented by some information from their Web sites.

Table 4.28 Non-Roman Scripts in Bibliographic Utilities

Features	RLIN*	OCLC**	Remarks
Scripts	JACKPHY (CJK-1983, Arabic/ Persian 1991, Hebrew/Yiddish-1988, and Cyrillic-1985.	CJK-1986 Arabic-2001	RLIN provides access to more scripts than OCLC.
Hardware	PC-based multiscript workstation, supporting CJK, Arabic, Hebrew/Yiddish, and Cyrillic.	PC–based Multipurpose workstation with WinNT 4.0, Win2000, Win. XP OS for CJK; Win2000 and Win. XP for Arabic vernacular scripts support. All roman-alphabet languages are supported.	No difference
Keyboard	Standard keyboard for CJK, Arabic, Hebrew, and Cyrillic scripts.	Standard, with virtual keyboard for Arabic scripts.	OCLC's use of Arabic virtual keyboard is useful in some settings.
Software	Interface with dBase III & Lotus 1-2-3	Full Windows-based application with online interactive and batch modes using offline local file; several host interface transactions can be done in batch processing; labels, accession lists; includes comprehensive Help.	OCLC's software configurations are more versatile than RLIN's
Input method	Character-components entry method is still an option (now mapped to a standard keyboard). RLIN also has offered Romanization (pronunciation)-based input methods for CJK since the introduction of "RLIN Terminal for Windows" software.	Four pronunciation-based and one character-component based entry methods., plus Auto-transliteration for Chinese pinyin fields entry based on Chinese character entry codes; Text string phrase bank with macro use and ALA diacritics and special character entry by table or macro keys for CJK and Arabic. Arabic includes an automated transliterator that transliterates Arabic romanized data into the Arabic script.	OCLC's Auto-transliteration will save time input and may increase accuracy.
Character sets	Incorporates all CJK characters into one character set. RLIN uses the same Arabic, Cyrillic and Hebrew characters that are defined in the MARC 21 sets.	ANSI Z39.64-1989 (East Asian Character Code) for CJK MARC storage in the Host and outputs. Unicode (2.1) for CJK record display since 1998. MARC 21 Basic and Extended Arabic character set for Arabic	Switching between character subsets creates additional keystrokes and breaks the consistency of typing
Search method	Word, phrase, Boolean operators, etc.	CJK vernacular personal and corporate names, title, and name/title, and combined name and title searches. For Arabic, personal name, corporate name, name/title, title, combined name and title.	RLIN offers much more powerful search capabilities
	Transmission by pages	By the entire record	No difference.

* Most of the information in this table that pertains to RLIN was kindly provided via email communication with Diana Hall, RLG Information Center.

** Most of the information in this table that pertains to OCLC was kindly provided via email communication with Hisako Kotaka, Senior Product Manager, Metadata Services Division, OCLC, Inc.

Thesaurus	Online reference to 35,000 CJK characters	Electronic CJK Dictionary with cross references for EACC, Unicode, fonts, graphic and phonetic input codes for CJK; several indexes and copy/paste available for E-dictionary use	OCLC's electronic dictionary provides more capabilities
Authority file display	Display all diacritics	No diacritics display on the list output; diacritics included on a single record display	OCLC's authority file search may result in confusion and misleading results
	Authority searches have to be done separately	Permits the authority file access while inputting; Save to and export authority records from the authority local file	OCLC's capability is necessary and convenient
Communication costs	Most users connect over the internet: $3.85/hour or $125/user/month for connect time.	Internet, TCP/IP access, Dial access.	Access costs are affordable for a single library
Solutions	The Web-based search interface Eureka®. Requires a Unicode-compliant Web browser and the appropriate repertoire of fonts.	Stand alone, Windows-based software "Multiscripts Z39.50 Client". Requires a Unicode font installed on the OPAC workstation (Arial Unicode MS font is recommended).	RLIN's Eureka seems more manageable.
Database size	As of Mar. 2002: - 1,325,574 unique titles (*not* records) with CJK scripts - 185,594 unique titles (*not* records) with Hebrew scripts	- 2,149,833 unique records with CJK scripts (July 1, 2002) - 35,311 unique records with Arabic scripts (October 1, 2002)	CJK records have a lower hit rate on the RLIN database than on OCLC's. OCLC & RLIN agreed to exchange CJK records created after September 1988 through LC

Comparing the two utilities, suggests that RLIN provides access to more non-Roman scripts and more advanced search capabilities. OCLC has an enhanced online CJK thesaurus. Since it has a larger CJK database than RLIN, searching and evaluation of the matching records on OCLC are less time-consuming with a higher hit rate. With regard to connection costs, using both of the utilities seems to be affordable for today's library budgets. Meanwhile, both of the utilities were able to take advantage of more advanced technology and offer some "state of the art" features. For instance, OCLC uses a virtual Arabic keyboard and automatic transliterator for Chinese pinyin. On the other hand, RLIN's Web-based searching interface, Eureka, is easy to set up and requires less time for the system administrator.

4.2 Public Services for Non-Roman Script Materials

The questionnaire instrument was designed to collect data with regard to public services, e. g., reference services, provided to facilitate use and access of the non-Roman script materials among the sample libraries. Items that were of interest to this study included librarians' and users' access to transliteration tables, the provision of bibliographic instruction or training, the availability of instructional materials, the availability of reference librarians with language skills, and finally the problems encountered by these librarians when helping users find and access non-Roman script materials. The next five sections provide the analyses.

4.2.1 Access to Transliteration Tables

In a totally romanized environment, or even where Romanization is used alongside with vernacular scripts, public service librarians' and end users' direct access to transliteration tables of the Romanization schemes is crucial. Figure 4.18 illustrates the percentages for each group of users.

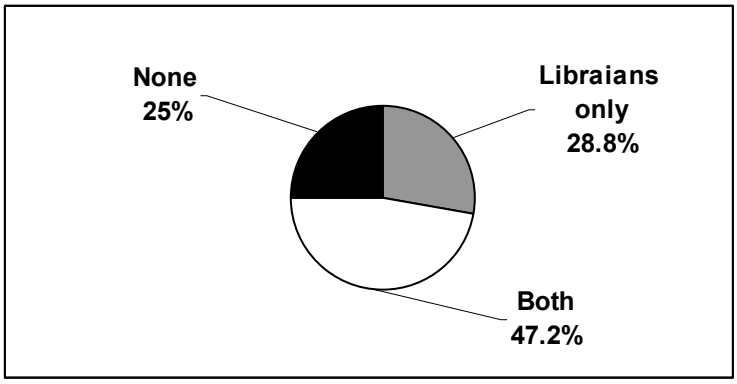

Figure 4.18 Access to Transliteration Tables

About half of the respondents indicated that both reference librarians and end users have access to transliteration tables (47.2%). In other words, almost half of the respondents (n=36) surveyed provide access to transliteration tables neither to reference librarians nor to end users. Another significant figure in this analysis is that about 53.8% of end users do not have any access to these tables.

4.2.2 The Provision of Bibliographic Instruction

Bibliographic instruction or training is a commitment in most academic/research libraries to ensure that end users have the required skills to easily identify and access the information they need for study, research, and teaching purposes. Figure 4.19 illustrates the provision of this important public service among respondents (n=37).

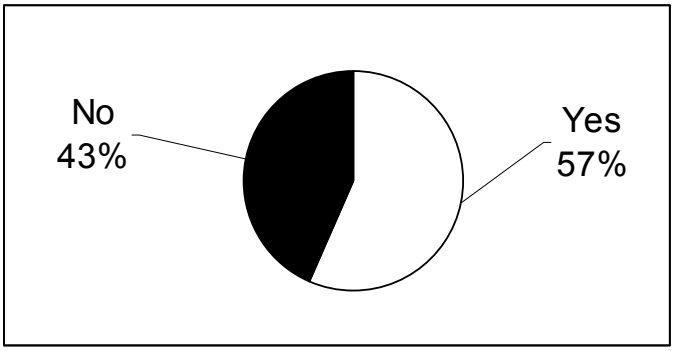

Figure 4.19 The Provision of Bibliographic Instruction

Up to 57% of the academic libraries surveyed indicated that they offer bibliographic instruction to end users on searching for non-Roman script materials in the library OPAC, regardless of whether Romanization and/or original (vernacular) script is used. Some libraries specified that subject specialists or bibliographers, not the reference librarians, give this instruction. Others indicated that such instruction is given only upon request. However, there are 43% of academic libraries that responded to the survey that do not provide any bibliographic instruction to their end users.

4.2.3 The Availability of Instructional Materials

Another aspect of bibliographic instruction as a public service is the provision of instructional/training materials, e.g., handouts, brochures, online help file within the library OPAC, electronic guides on the library's Web site, etc. Figure 4.20 illustrates how far academic libraries surveyed go for providing these materials.

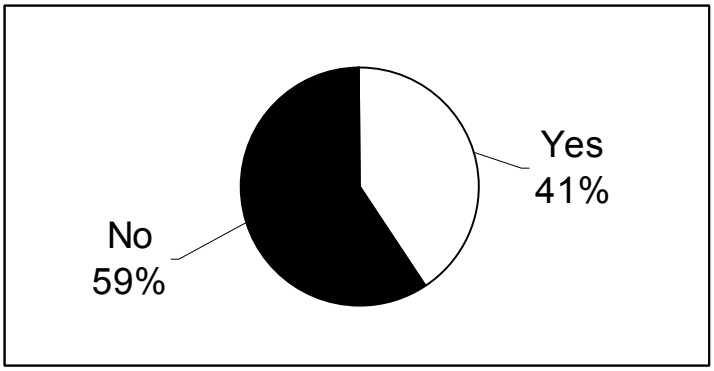

Figure 4.20 The Availability of Instructional Materials

Respondents to this question (n=37) were divided, with a little above 40% providing instructional materials to help find non-Roman script materials in the OPAC, and nearly 60% having no such materials prepared for their users.

4.2.4 The Availability of Public Services Librarians

The human aspect of any public service in academic/research libraries may be measured by characteristics of professional librarians, such as work experience, technological competencies, and personal and communications skills, etc. Another essential aspect of these characteristics, especially in libraries with sizeable non-Roman script materials, is the availability of public service librarians who posses language skills to better assist patrons who are interested in these materials. These skills could be classified as part of the personal and communication skills and will include not only the ability to speak a particular language, but also the ability to write and

understand the written script. Table 4.29 represents the frequency distributions and percentages of the responses to this item of the questionnaire.

Table 4.29 The Availability of Public Service Librarians with Language Skills

Public Service Librarians	Yes (%)	No (%)	Totals (%)
CJK Librarians	23 (62.2%)	14 (37.8%)	37 (100%)
Middle East Librarians	16 (38.2%)	21 (56.8%)	37 (100%)
Cyrillic Librarians	16 (43.2%)	21 (56.8%)	37 (100%)

The availability of public services librarians with CJK language skills is as high as the availability of these collections in the sample libraries, with 62.2% having librarians with CJK language skills. The figures of Middle East and Cyrillic public service librarians are exactly the same. Among the respondents of each group, 38.2% indicated that they have enough librarians with these language skills.

4.2.5 Romanization Problems as Encountered by Public Service Librarians

The last item in the public services part of the questionnaire instrument requested input from respondents with regard to the problems encountered by the reference librarians concerning Romanization. Figure 4.21 illustrates the frequency distributions of these problems.

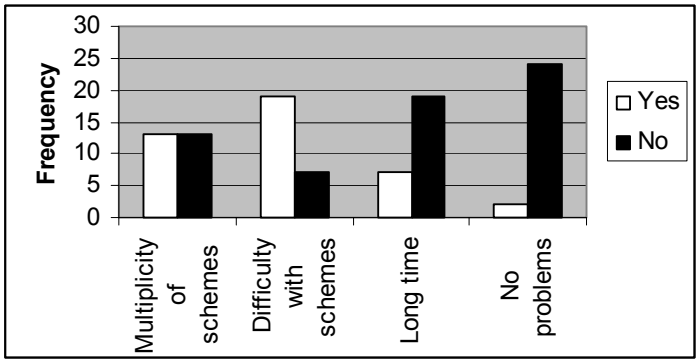

Figure 4.21 Public Service Librarians' Romanization Problems

The first observation of the bar chart above is that 92.3% of the respondents (n=26) encounter problems with Romanization when helping patrons find non-Roman script materials. The second significant observation is that nearly three quarters of these respondents experience difficulties when applying/using the Romanization schemes. That different schemes need to be consulted when searching for non-Roman script materials is indicated as a problem by 50%. And finally, the long amount of time public service librarians spend when searching for non-Roman script materials was regarded as a moderate problem by 26.9% of the respondents.

Here are some other comments from the respondents, in their own words:

"Frustration on part of patron: for students of language, they have to learn another scheme to find items they want; for native speakers may be familiar with a non-L.C. Romanization and/or also frustrated because they can search in their home country in their language; finally, they are frustrated because they can search the Web using their own non-Roman script language, and they don't understand why they can't in the library."

"Users are accustomed to seeing so many non-systematic Romanization [schemes] that they don't often request help in learning which system this library uses."

"It has been my experience (for 25+ years) that library staff who do not know the non-Roman script don't even bother to look at romanized info – they throw up their hands and say: "I can't read this"! This is especially true of non-professional staff in public services."

4.3 Profiles of Automated Library Systems

Automated library systems became an integral part of the provision of bibliographic access to library collections. These systems have worked, and continue to work for Roman script collections since the early days of library automation. However, with the widespread use of non-Roman script collections in academic/research libraries, the efficiency of these systems to support the bibliographic access to vernacular scripts needs to be reexamined. For that reason, the questionnaire instrument contained one part that examined the automated library systems from the point of view of support for non-Roman scripts.

4.3.1 Automated Library Systems

This section provides the results of responses to the first question in the "Automated Library System" part of the questionnaire. Respondents were asked to specify which integrated library systems they use for managing library collections. Table 4.30 represents the frequency distributions and percentages of their responses (n=39).

Table 4.30 Automated Library Systems in Academic Libraries Surveyed

Systems	Frequency	Percent	Cumulative Percent
Innopac (Millennium)	13	33.3	33.3
Voyager	5	12.8	46.2
Aleph	4	10.3	56.4
Unicorn	4	10.3	66.7
NOTIS	4	10.3	76.9
Horizon	3	7.7	84.6
VTLS	2	5.1	89.7
Taos	1	2.6	92.3
DRA	1	2.6	94.9
Advance	1	2.6	97.4
Locally developed	1	2.6	100
Total	39	100	

There are a total of 10 commercial systems, plus one locally developed system, being used in the sample libraries. Based on Table 4.30, these libraries could be divided into four categories.

1. The largest group of libraries surveyed, 33.3%, uses Innopac (Millennium).
2. The second largest group of libraries, 12.8%, uses Voyager.
3. Three different systems are equally used in a total of 12 libraries surveyed, or 33.3% of the respondents: Aleph, Unicorn, and NOTIS.
4. Four different systems are used each in four different libraries, or 10.4%, of the respondents: Taos, DRA, Advance, and a locally developed system.

These automated systems are being used mostly in the standard 32-bit Windows environment; such as Windows 9x, 2000, XP, NT in staff and in end users' devices as well. Few libraries indicated the use of Mac operating systems. Most of the respondents pointed out that their OPACs are Web-based, meaning that the version the browser is also important in order to smooth the search and display of vernacular characters in the bibliographic records.

4.3.2 Consideration of Scripts Support upon System Implementation

Respondents were asked to identify whether they have considered the support for non-Roman scripts as an important functionality. Figure 4.22 below illustrates that 41.7% of the respondents (n=36) did consider such support as an important functionality. Some other libraries stated that, "While non-Roman support in future releases was discussed, it was not a significant determinant in system selection."

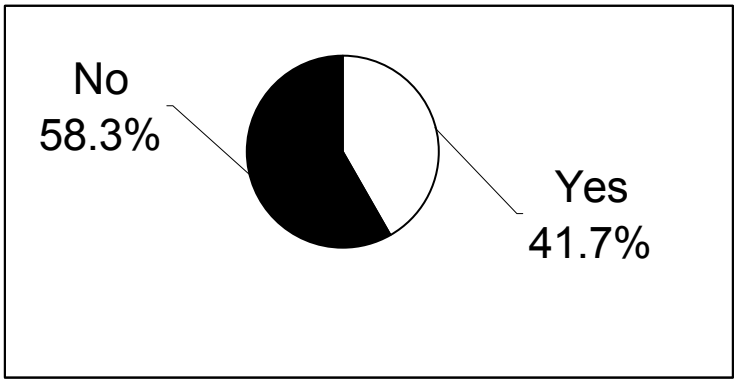

Figure 4.22 Consideration of Vernacular Support upon Implementing Automated Systems

Some self-explanatory comments from the respondents are worth noting here:

- [Script support] feature requested by faculty members in East Asian Studies.
- Although this wasn't a feature 13 years ago, we anticipated that the vendor would provide it in the future, by implementing Unicode.
- We had over 40,000 CJK records which we wanted to integrate with the ILS.
- We listed it as a "desirable" feature, but at the time it was not generally available, so was not a make-or-break feature.
- Our library has a longstanding commitment to build collections in East Asian languages. Although full functionality for CJK has at times seemed more like a dream, we have always hoped and expected our automation vendor to deal with this issue.
- When current system specification written, attention was paid to non-Roman scripts. Only limited support available in 1993. We are in the process of implementing full system support to non-Roman characters. Expect project completion before the end of this (2002) year.

4.3.3 Future Consideration Regarding Script Support

The 21 respondents who said "No" to the item discussed in 4.3.2 were asked to identify whether the support for non-Roman scripts in library OPACs will be considered in the near future (within 5 years). Figure 4.23 shows the percentages of their responses.

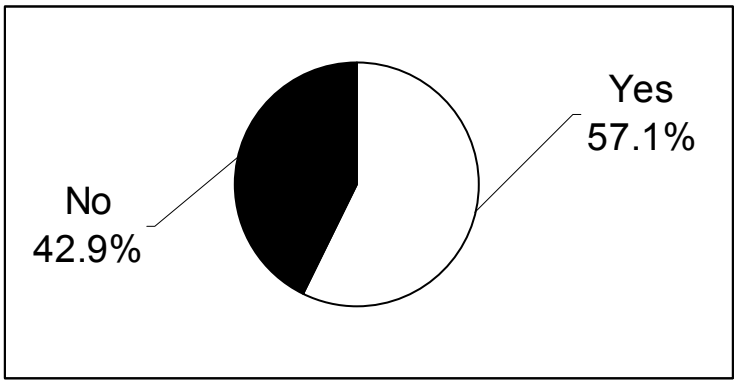

Figure 4.23 Future Consideration Regarding Script Support

While 42.9% of the respondents will not consider vernacular scripts support in the near future because of the potential cost and use of non-Roman alphabets, there are 57.1% who indicated such consideration. Below are some of their comments in this regard:

- We will migrate to a new system in 2003. We considered supporting non-Roman scripts as important functionality.
- "Essential" may be a bit strong, but we definitely will consider its improvement as very important.
- We are in the middle of a process to select a new integrated system and vernacular capabilities will be a consideration.

Table 4.31 represents the crosstabulation of both catalogers' and systems librarians' considerations toward vernacular support in the near future. Looking at the CJK column, there is 50% agreement to consider such support versus 50% disagreement or confusion that such support will be considered in the near future.

Table 4.31 Crosstabulating Catalogers' and Systems Librarians' Future Consideration Toward Vernacular Support

Systems Librarians	CJK Catalogers (n=4)		Cyrillic Catalogers (n=11)		Middle East Catalogers (n=7)	
	Yes	No	Yes	No	Yes	No
Yes	1 (25.0%)	1 (25.0%)	5 (45.4%)	3 (27.3%)	4 (57.1%)	2 (28.6%)
No	1 (25.0%)	1 (25.0%)	0	3 (27.3%)	0	1 (14.3%)
Total (%)	2 (50.0%)	2 (50.0%)	5 (45.4%)	6 (54.6%)	4 (57.1%)	3 (42.9%)

In the middle of the table there are eight libraries with Cyrillic that showed a complete agreement between catalogers and systems libraries in this regard; while three other libraries have a conflict between these two groups, as systems librarians are considering vernacular support that is not expected by catalogers.

Finally, the far right column demonstrates that five libraries with Middle East script materials or 71.4% of the respondents have an agreement about future plans regarding vernacular support between catalogers and systems librarians; while 28.6% of these libraries have some confusion.

4.3.4 Profiles of Automation Vendors and their Systems

This section provides an overview of each of the vendors whose systems have been selected by members of the Association of Research Libraries that responded to the survey. According to Breeding (2002), "The members of the Association of Research Libraries represent the high end of the library automation marketplace.... The systems selected by these libraries can be considered the best available for large research libraries."

General Comments:

- Library automation vendors recognize the importance of Unicode support in order to find broader markets for their products. As some vendors indicated that they are moving toward upgrading their systems to implement the Unicode standard with 2-3 years, one can expect that this standard will be a strategic component in the major automated library systems by the year 2005.
- Costs: All vendors supporting non-Roman characters indicated no additional costs to get their systems with these capabilities. This may be because Unicode support is inherent throughout all aspects of any software supporting non-Roman scripts. Even vendors who plan to implement Unicode support within 2-3 years indicated no additional costs to the customers, as one vendor stated, "*When we add the support, we do not plan to charge our libraries additionally*."
- Special (or extra) hard/software: All vendors indicated that there is no need for any additional hardware. The only piece of hardware that may be needed is a keyboard with the language particulars. However, all vendors indicated the need for either a copy of the operating systems (e.g., Windows) that is designed specifically for certain language, or Windows 2000 (or higher) for full Unicode support.

Innovative Interfaces Inc. (III)
System: Innopac (Millennium)
URL: www.iii.com

In a press release dated June 15, 2002, III announced that members of the Association of Research Libraries (ARL) continue to embrace Millennium as their library automation solution. Innovative's high number of ARL members prompted Library Journal to note, "Innovative's 34 ARL library customers continue to give it the highest penetration into ARL's 122 institutions."

With regard to non-Roman support, Innopac (Millennium) has full integrated Unicode, and supports cataloging in a wide range of languages and character sets (Chinese, Thai, Japanese, Korean, Hebrew, Cyrillic, and Arabic). While the U.S. MARC format specifies that non-Roman characters appear in parallel 880 fields, Innopac (Millennium) goes beyond that standard. "*The basic design of INNOPAC Millennium allows non-Roman characters to appear in any field, in any record*," according to Steve Silberstein, Innovative's executive vice president. III also provides "Internationalization and localization" as part of the Innopac (Millennium) system offerings.

Respondents' experiences with Innopac (Millennium):

- To display CJK in all OPAC workstation, use CJK Module, and set the encoding to Unicode UTF-8 on Internet Explorer to view CJK characters.
- There is a separate Web OPAC Port for searching and displaying CJK records. It is directly linked from Web OPAC opening pages.
- The original scripts are fully supported by Unicode, but keyword searching is not capable.
- The vendor offers the option of multilingual interfaces; however, our library did not purchase the feature.
- To search and display CJK in special workstations, Use IE and Microsoft IME JCK. Records in Unicode display.
- III also supports Arabic/Hebrew script capabilities.
- Through IME capabilities on Windows 2000, the library can offer patrons ability to enter searches and view text in CJK languages.

Endeavor Information Systems
System: Voyager
URL: www.endinfosys.com

In a press release dated Aug. 13, 2000, Endeavor introduced an enhancement to the Voyager integrated library management system through Unicode capabilities. That was followed by the establishment of the "Unicode Task Force," which includes members from Cambridge University, Getty Research Institute, University of Hawaii, Library of Congress, Linnea2 Consortium of Finland, University of Pittsburgh, Pepperdine University, Princeton University and Yale University.

Work is in progress to implement Unicode fully into the Voyager bibliographic (and authority) structures. This project is being conducted in several phases: display of non-Roman scripts (completed); keyboard input and editing (planned for 2002 release); sorting and searching; and localization (no firm schedule of completion).

Currently, Endeavor's non-Roman scripts support is based on a special product "Glyph Server," developed in partnership with InterPro Global Partners, and converts MARC data into Unicode, then displays the characters in WebVoyage (Voyager's OPAC module) as images of the language-specific glyph set designated in the MARC record. This means that these characters will not be searched by end users. Searching has to be done in Romanization. However, there are some advantages for using the Glyph Server. First, Libraries don't need to worry about loading huge (and often incomplete) font sets onto library and remote PCs; second, the display of Latin 1 characters isn't comprised by conflicting font sets; and finally, Punctuation appears in the appropriate spots for right-to-left languages. The Glyph Server supports all of the JACKPHY languages (Japanese, Arabic, Chinese, Korean, Persian, Hebrew and Yiddish), as well as a number of other scripts.

ExLibris USA, Inc.
System: Aleph
URL: www.exlibris.com

ALEPH 500 (starting version 14.2) supports the UTF-8 (Universal character set Transformation Format, 8-bit), frequently known as "Unicode." The library may enter, edit and view the data in different ways using ALEPH character conversion tables.

Currently all the non-Roman languages of interest to this study, in addition to many others, are supported by the ALEPH 500 system: Russian, Hebrew, Arabic, Chinese, Japanese, and Korean. ALEPH supports the direct input and storage (in a convertible but proprietary format) of Cyrillic, Greek, Hebrew, Arabic and CJK (Chinese, Japanese, and Korean) characters.

OPAC screens can be offered in up to 21 languages, and additionally, patrons can choose – by a session preference or a permanent profile – whether to limit the retrieved records to a particular language. The Microsoft floating (virtual) keyboard enables the user to enter non-Roman scripts through the keyboard (these characters must be present in the PC character set). The floating keyboard is simply a table-driven set of buttons with texts and ASCII values associated with these buttons.

Respondents' experiences with Aleph:

- ALEPH 500 system started in 14.2 version has Unicode supported CJK display functionality, 15.4 has both display and search functionality, and good editing functionality in CJK.
- The system will display Cyrillic in bib. Records, but we don't [want to].
- ALEPH theoretically supports vernacular but we are in early stages of implementation and have imperfect display, no indexing, and no input capability.
- We are working now to figure out proper display of the characters in parallel fields that are indexed and searchable.

Sirsi, Inc.
Systems: Unicorn, Taos, DRA
URL: www.sirsi.com

In a press release in May 17, 2001, Sirsi and Data Research Associates, Inc. (DRA) jointly announced the signing of a definitive merger agreement. One year later, more than 60 DRA Classic, Inlex, Taos, and MultiLIS sites have chosen to upgrade to the Sirsi's Unicorn Library Management System.

Sirsi offers a WebCat native language interface in Arabic, Chinese, Spanish, German, and French. With one click of the drop-down toolbar, users may select the language in which they want to search library holdings and access library resources. The entire screen, including toolbars and buttons, is immediately displayed in the language of choice. Libraries also have the option of providing one particular language at a selected workstation or workstations. The latest released version, 1.4.1e, supports spell checking and display of Arabic, Cyrillic, and Hebrew character sets.

Epixtech, Inc.
Systems: Horizon, NOTIS
URL: www.epixtech.com

Epixtech, formerly Ameritech Library Services, became a private corporation in December of 1999. The epixtech academic solution has its roots in the NOTIS software. NOTIS was the system of choice for the majority of ARL libraries in the 1980's and 1990's. Today, Horizon Sunrise version 7.2 modules are Unicode enabled, providing libraries with international support. The use of Unicode extends diacritic input and editing from Latin 1 (Roman) to non-Roman languages.

Horizon Sunrise's major features and benefits include: support of ALA diacritics and Unicode for both display and editing; support of diverse languages: English, CJK, Vietnamese, Spanish, and Greek; diacritics and special characters will display in all records, not just MARC records; keyboard or graphical input methods; ability to print spine labels with diacritics; and following Unicode conventions for the input of diacritics and special characters. In addition, in the database level, Horizon Sunrise 7.2 can store Unicode characters in all tables; has stronger multilingual support; and allows for correct sorting sequence.

Respondents' experiences with Horizon:

- (CJK-Display only): searching is possible. We are testing the feasibility; have not implemented.
- (Cyrillic-Display only): We can display the characters if they are in the records. We are testing the indexing and searching capabilities.

Respondents' experiences with NOTIS:

We use a highly modified version of NOTIS. Our system can display both Romanization and vernacular, but patrons can search only by Romanization, not vernacular. Searching by vernacular is on next upgrade. We use Unicode to display our Web-based catalog CJK citations on Internet Explorer v. 6.

Geac, Inc.

System: Advance
URL: www.geac.com

After thorough examination of the vendor's Web site and the Web in general, in addition to library literature databases, and even a contact with the vendor via a Web request form, I could not find any information related to the non-Roman script support in Geac's ADVANCE system. The assumption is that the system does not support non-Roman scripts in bibliographic records.

Respondents' experiences with Advance:

- We build Arabic/Hebrew records in RLIN. Geac Advance does not have vernacular input/edit/display.

VTLS, Inc.
System: VTLS Classic
URL: www.vtls.com

VTLS does not generally sell VTLS Classic system anymore as their Virtua system is the most advanced technological system the company developed. VTLS Classic used to support non-Roman scripts in the library OPAC and there are several customers who continue to utilize the Classic system worldwide. Many, however, have already migrated to the Virtua system. "We support over 200 sites worldwide for our Classic system, many of which utilize non-Roman scripts," said Dawn Stoneking-Thomas, Information Officer, Marketing, VTLS Inc.

The multi-lingual capabilities of VTLS Classic allows the patron several language options simultaneously and allows one to set the prompts in any language on any screen.[*]

VTLS will share some of its expertise in globalization with OCLC as they increase cooperative efforts in extending and enriching the OCLC WorldCat database. As part of an agreement signed March 8, 2002, OCLC will license source code for specific parts of the new Virtua ILS Integrated Library Systems server software developed by VTLS (Dorman, 2002). "VTLS has developed robust implementations in areas such as Unicode, Z39.50, thesaurus and authorities," said Lynn Kellar, director, OCLC Enterprise Database Technologies and project manager for extending the WorldCat database. "The work they have already completed will help us to accelerate construction of fundamental pieces of the extended WorldCat database."

[*] Email communication (10/21/2002) with Dawn Stoneking-Thomas, Information Officer, Marketing, VTLS Inc.

Virtua ILS (Integrated Library System) complies with the Unicode standard on both the client and server levels. The system stores data natively in Unicode. Users can import, catalog and display records in any language. Users can also change interface languages on-the-fly and store multiple scripts within the same MARC tag or sub-field.

Summing Up

The above profiles of automation vendors and their integrated library systems primarily demonstrate that library automation industry is moving toward a multilingual support that subsequently will provide a sound ground for bibliographic access to original (vernacular) scripts in library OPACs.

Recognition for the Unicode standard indicated that it will be a strategic component in the near future in widely used automated systems, like most of the ones currently implemented in academic and research libraries that responded to this survey.

However, the study results indicated that most libraries have difficulty configuring these systems up to their maximum capabilities. While displaying script characters is slightly supported, searching still needs extra work between vendors and their customer libraries.

CHAPTER FIVE

DISCUSSIONS AND IMPLICATIONS

5.1 Summary of the Research

Providing bibliographic access to non-Roman script materials has been a daunting task in academic and research libraries with sizeable collections of these materials. While these libraries have gained many advantages brought by automation, non-Roman materials created some technical difficulties in data processing. Throughout the last two decades, academic libraries have been trying to overcome these difficulties, along with two major developments that gave the opportunity to these libraries to facilitate bibliographic access to vernacular characters in their OPACs. These two developments are the increasing number of bibliographic records with script characters available from bibliographic utilities, and the development of a universal character set, the Unicode standard, that made processing the world's non-Roman characters feasible in the computer.

This study was designed to trace the effect of these developments on local practices of academic library members in the Association of Research Libraries (ARL). The purpose was three-fold as follows: first, to determine the characteristics of non-Roman script materials and how they are being cataloged (romanized vs. script cataloging); second, to identify the status of current OPAC systems in libraries where these materials are collected with regard to support of non-Roman characters (e.g., searching, displaying, etc.); and finally, to develop a set of functional requirements for OPAC designers and developers, as well as for librarians, concerning the desired features that should be available in library OPACs to facilitate the bibliographic access to these materials.

A purposive sample of 70 academic library members in ARL with sizeable non-Roman script materials was chosen after thorough examination of their Web sites and checking the members' directories of national associations whose membership is based on the availability of non-Roman collections. A self-administered questionnaire was mailed to each library in the sample, and resulted in a 69% response rate. However, the data analysis was based on 45 responses (65%) obtained by the cutoff date of the data collection.

In addition to gathering data from libraries, it was also necessary to contact the two major bibliographic utilities, OCLC and RLIN, as well as library automation vendors to obtain information with regard to their efforts in the area of vernacular support of bibliographic records. A document/Website analysis of these two stakeholders was also utilized to gather the needed data for this study.

Library questionnaires were analyzed and coded in the statistical package SPSS version 10.1 for Windows. Frequency distributions were obtained for each variable. Crosstabulation, a statistical technique for bivariate analysis, was used wherever the need arises to measure the association between two variables.

5.2 Discussion of Findings

Research Question One. The first question of interest to this study read, "What are the characteristics of non-Roman script materials of libraries from the population? How are these materials cataloged, stored, etc.?" Part one of the questionnaire, the collection profile, consisted of items that addressed this research question. Below are the major findings along with a discussion of their implications and particularly their relation to previous research.

The questionnaire data revealed that among the six non-Roman script collections studied, the three East Asian script materials ranked the first in terms of the availability of script

materials, followed by Cyrillic collections, and then came the Middle East (Arabic and Hebrew) collections. The considerable interest between the United States and East Asian countries and also Russia, especially after World War II and the cold war era, gave importance to collecting materials about the history, economics, politics, culture, etc. In the meantime, the widespread of educational programs and departments of languages and area studies made it necessary to develop collections that support the study and research in these disciplines.

With regard to shelving, East Asian collections tended to separate from the Roman collections. The Middle East collections (Arabic and Hebrew) were mostly integrated with Roman collections, and the Cyrillic materials are all integrated with the Roman collections. In other words, East Asian collections are the script materials that receive the most special consideration in academic libraries, in terms of devoting a special unit or department in the library with its own librarians who master the appropriate language skills to help patrons access and use these collections.

With regard to source of bibliographic records, East Asian collections relied on OCLC twice as much as on RLIN. Similarly, the Middle East collections tended to use OCLC more than RLIN. But the significant difference here was that a substantial portion of these collections use both of the utilities as a source of bibliographic records (35-42%). Although it does not focus on Cyrillic script cataloging, OCLC was used by 65.4% of the respondents to the Cyrillic part of the questionnaire as the sole source for bibliographic records; only nine out of 26 libraries used RLIN to catalog their Cyrillic collections.

The last finding raises a question, especially as we know that RLIN has been making available bibliographic records with Cyrillic characters since 1985. However, two other findings that were discussed in chapter 4 explain the situation. The first is the prevalent use of

Romanization for Cyrillic cataloging, and the second is the limited support for Cyrillic script in automated systems in academic libraries with Cyrillic collections (81% of these systems do not support typing in Cyrillic, and 84% cannot handle searching and displaying Cyrillic vernaculars.)

The library OPAC in academic libraries became the main gateway for accessing bibliographic records of non-Roman script materials. Very few libraries indicated that some unconverted records are still in a paper file, e.g., Card catalog. This finding suggests an end to the attitude in the mid nineteen eighties when Rogers (1985) pointed out that, "Until the day when technology has advanced to the point that bibliographical records with CJK characters can be inputted [sic] and retrieved as easily as one can type "abc" on the terminal keyboard without requiring special or expensive equipment, cards with CJK scripts will still be quite essential to accurate bibliographical identification."

The majority of respondents with CJK collections performed copy and original cataloging in the combined romanized and vernacular characters. On the other hand, the survey results indicated that the majority of Cyrillic cataloging was performed using Romanization. This dominant romanized-based bibliographic access for Cyrillic collections should raise a question with regard to end users' abilities to successfully retrieve relevant items. Previous research concluded that romanized-based Cyrillic retrieval bears many deficiencies (see for example, Pasterczyk 1985, and Aissing 1995). This could be interpreted in light of the widespread use of OCLC, which doesn't not support Cyrillic vernacular cataloging. RLIN does support Cyrillic, but is less widely used. Also, the major automated systems as currently used by these libraries do not support Cyrillic characters.

In the case of Arabic script materials, 52.6% obtained romanized cataloging vs. 47.4% who used a combination of romanized and vernacular Arabic scripts in bibliographic records.

The situation was different in Hebrew cataloging where 75% of the respondents (n=20) performed romanized cataloging vs. 25% for a combination of romanized and vernacular Hebrew scripts in bibliographic records.

With regard to reasons behind cataloging non-Roman script materials in original (vernacular) characters, the study results revealed that providing better access to users of the collection was the highest selected reason for cataloging in vernacular, followed by taking advantage of the bibliographic utilities as they started to provide access to bibliographic records in the vernacular. Having librarians with the required language skills was also a determining factor (50%) that encouraged libraries with CJK collections to catalog in the vernacular.

Research Question Two. The second research question asked, "What is the current status of non-Roman scripts in library OPACs in these libraries? What are their plans for the near future (within five years)?" Part two of the questionnaire, the OPAC profile, addressed this research question. In addition to part two, the last part of the questionnaire, systems department, provided data about automated library systems in libraries from the sample. Below are the major findings along with a discussion of their implications and particularly their relation to previous research.

With regard to the interface language of automated library systems, almost none of the automated systems in academic libraries that responded to the survey support the capability to change the interface language to users' preferences. Although the interface language may not have any effect on the search and display of vernacular characters, some authors have emphasized its importance in relation to user-centered or user oriented design (Tyckoson 1991, Chepesiuk 1997, and Babu and O'Brien 2000).

The systems' vernacular typing capabilities varied for each of the three groups of vernacular scripts. The support for entering CJK scripts in cataloging modules was the highest, with 42.9% for Chinese, 44.4% for Japanese, and 40% for Korean. The capability to enter Arabic vernaculars when cataloging Arabic script materials was fourth in order (25%). About one fourth of the respondents determined that their automated systems support the entering of Cyrillic script (19%). And finally, 14.3% of the respondents indicated that they could enter Hebrew characters in the cataloging modules of their automated systems.

While the majority of original cataloging is performed in the bibliographic utilities' system/interface, the capability to type in non-Roman vernaculars is still necessary for purposes of changing downloaded records to conform to practices of the local library system. The lack of such capability may hinder libraries from using their automated systems up to their maximum benefits, and force librarians to use bibliographic utilities as the cataloging interface and the local automated system as the display-only interface.

Concerning automated systems' capability to search/display scripts in library OPACs, about 20% of the respondents indicated that they can search and display in Chinese and Japanese, while 15.4% indicated this capability in Korean. Approximately 30% of the respondents affirmed the capability to "display only" in their automated systems the three East Asian scripts. The majority of respondents indicated that end users cannot search and display Cyrillic characters in the OPACs, while the capability to only display vernacular Cyrillic constitutes 12%, and the ideal, searching and displaying, represents just about 4%. Furthermore, 73.7% and 85% of the respondents to the Middle East part of the questionnaire indicated that their OPACs do not support the search and display of Arabic and Hebrew, respectively. About 10% of the OPACs in these libraries can display Arabic/Hebrew in the bibliographic records.

These findings reflect the dilemma of providing bibliographic access in library OPACs for vernacular characters already downloaded from the utilities. For example, two thirds of the libraries that obtain Chinese copy cataloging with vernacular characters, and about one third of libraries that obtain Arabic copy cataloging with vernacular characters cannot allow their end users to search and display these records in vernacular due to limitations of automated systems.

Most libraries surveyed indicated they already have the records in their databases, but the script characters are not searchable or displayed. The system plans here showed that these libraries are either upgrading to new systems with the capability to support non-Roman characters or testing their current systems that already have this capability, but libraries have technical difficulties configuring these systems to handle vernacular characters in bibliographic records.

Libraries with romanized CJK bibliographic records (n=4) have plans for the near future, within 5 years, considering vernacular cataloging (50.0%) vs. 50.0% who do not. Although they catalog their collections only in Romanization, only 43.5% of the respondents to the Cyrillic questionnaire indicated future plans for vernacular cataloging. The majority of respondents to the Middle East questionnaire (n=15) indicated that they have future plans for cataloging Arabic and Hebrew script materials in their vernacular (73.3%).

There are a number of possibilities for the low percentage of future plans regarding Cyrillic cataloging presented above: (1) two thirds of Cyrillic catalogers indicated that Romanization is adequate for their library goals, (2) two thirds of the automated systems in libraries with Cyrillic collection do not support Cyrillic vernaculars, and (3) two thirds of libraries with Cyrillic collections use only OCLC for Cyrillic cataloging, while there is no

support in OCLC for Cyrillic vernaculars. However, 58.3% of Cyrillic catalogers agree that cataloging of Cyrillic materials should be in their vernacular script.

It was determined that two thirds (68.8%) of libraries with CJK collections provide access to CJK records in all workstations throughout the library. The rest of the libraries (31.2%) have special devices allotted to provide access to these scripts in particular. The situation is different in Arabic/Hebrew access devices; one half (50%) of these libraries provide access to Arabic/Hebrew vernacular records in special workstations and the other half (50%) provides that access in all OPAC workstations in the library.

Finally, among the systems librarians responding to the questionnaire, 57.1% indicated that their libraries have future plans considering vernacular support in the OPAC. However, the bivariate analysis for system librarians and catalogers of script materials revealed that some libraries have contrasting visions between what catalogers are expecting and what system librarians are planning. For example, the disagreement between CJK catalogers and systems librarians is 50%, and reaches 28.6% among Middle East and 27.3% among Cyrillic catalogers.

Better coordination and continuing communication among non-Roman script catalogers and systems librarians are required to assure more accurate future planning. Otherwise, systems librarians may bring in systems that will not be totally accepted or equipped to support catalogers' plans regarding the bibliographic control of these materials. The proper involvement of catalogers in the creation of requests for proposals (RFPs) of library automation systems provides a higher potential to similar visions toward the implementation of the ideal automated system for library goals and plans.

The document/Web site analyses of library automation vendors revealed that some vendors have complete support for non-Roman characters, while work is in progress by other

vendors to fully implement the Unicode standard into their automated library systems in the near future. These developments could be further enhanced by incorporating input from both library catalogers and systems librarians with regard to the desired system that will fit not only present library needs but future ones as well.

Research Question Three. The third research question was, "What are the problems facing librarians, catalogers, and public service librarians, with regard to Romanization? And what are the obstacles that hinder vernacular access?" Some items throughout the questionnaire dealt with this research question, e.g., questions 8, 10, 17, 24 in each of the three script parts of the questionnaire, and question 5 in the public services part. Below are the major findings along with a discussion of their implications and particularly their relation to previous research.

Three main reasons for romanizing non-Roman script materials were identified by the respondents: the lack of automated system support for vernacular characters, the high cost associated with vernacular cataloging, and the shortage of staff with appropriate language skills. Romanization was considered adequate by 73.1% of Cyrillic catalogers. Because this position of Cyrillic catalogers contrasts with CJK catalogers, who indicated that Romanization is inadequate at all and also with Arabic/Hebrew catalogers, among whom just 27.8% who indicated adequacy of Romanization, I spoke with Brenda Carter, Slavic Languages Team Leader, University of Pittsburgh Libraries, who explained, "It may be that they believe that users are familiar with the [ALA/LC] transliteration scheme, or if they are not, it is relatively easy to learn. Also, until their bibliographic utility and/or their OPAC can allow for the display and searching of vernacular, it is not worth doing." However, we also learned that 58.3% of Cyrillic catalogers agreed that cataloging of Cyrillic materials should be in their vernacular characters.

The possible obstacles given in the questionnaire that hinder vernacular cataloging were in the following order: (1) No support in the automated system, (2) It may be costly, (3) Difficulties when obtaining the records with scripts from bibliographic utilities, (4) The current staff lack the appropriate language skills, and finally (5) No barrier, Romanization is considered adequate. As was expected, the lack of automated system support for vernacular characters was a major barrier in providing bibliographic access to non-Roman script materials.

Catalogers of the East Asian collections experienced some difficulties with Romanization: 33.3% of them had difficulties in applying/using Romanization scheme(s). The long time it takes to catalog a book was a special problem encountered by 33.3%. Cyrillic catalogers were the least among the three groups to experience any problem with Romanization. Middle East script catalogers were the most likely to experience problems with Romanization (73.3%). Among these problems were the difficulties in applying/using the scheme(s) (57.9%), and the long time it takes to catalog a book (52.6%). As the majority of libraries surveyed have catalogers with the appropriate language skills (see Table 4.23), these problems could be eliminated or at least reduced by providing access to script characters in bibliographic records.

The characteristics of Cyrillic, Arabic, and Hebrew scripts might impact the efficacy of their Romanization. First of all, the number of characters in each of these scripts is different than number of English Roman characters. While Hebrew has 22 and Arabic has 28 characters, Cyrillic script has between 30 and 40 characters. This means that there cannot be a one-to-one matching when romanizing bibliographic data. Second, the direction of Cyrillic writing is same as in English, but both Arabic and Hebrew are written from right-to-left. Third, in terms of context, Cyrillic characters are not sensitive, e.g., the character shape is the same wherever its location in the word. However, Arabic characters are context-sensitive. And finally, both Cyrillic

and Arabic make use of diacritics, requiring a cataloger who is familiar with the script to be able to romanize it correctly. In my interview with Brenda Carter, Slavic Languages Team Leader, University of Pittsburgh Libraries, she indicated that "ALA/LC transliteration for Cyrillic is relatively easy to learn. Cyrillic characters are neatly represented by the Roman equivalents. My impressions are that Hebrew/Arabic transliteration is much more difficult."

With regard to problems encountered by reference librarians with Romanization, 92.3% of the respondents encountered problems when helping patrons find script materials, as follows: 75% of the reference librarians surveyed experience difficulties when using Romanization schemes; the different schemes that need to be consulted were described as a problem by 50%; and finally, the long time spent when searching for non-Roman script materials was regarded as a moderate problem by 26.9%. These problems will continue to exist, even with the availability of vernacular characters in library OPACs, in light of the shortage of public service librarians with the appropriate language skills in academic libraries surveyed (see Table 4.29).

5.3 Functional Requirements for Non-Roman Scripts in Library OPACs

This section deals with the fourth research question that asked, "What are the functional requirements that could be recommended to designers and developers of library OPACs, as well as librarians, with regard to the support of vernacular characters?" The following functional requirements have been extracted from the questionnaire responses, vendors' brochures, and the in-depth analysis of the related literature.

These requirements have been developed specifically to minimize communication problems between system managers in academic libraries and automation vendors. The lack of understanding of these functional requirements hinders OPAC designers and system managers from making the best decisions when designing and managing their systems.

- The system must accept, load and store bibliographic records containing 880 fields (Alternate Graphic Representation), and store multiple scripts within the same MARC tag or sub-field. These records may be included in a library's initial data migration, batch-loaded or loaded online. The survey results revealed that some libraries catalog in the vernacular but do not have the appropriate support in their automated systems to allow searching and displaying. For example, among the respondents who obtain Chinese copy cataloging in vernacular, 66.7% cannot search and display these characters in the library OPAC.
- The system must allow loading bibliographic records with scripts from RLIN or OCLC. Bibliographic utilities have their own settings or requirements that local systems have to adhere to in order to work smoothly with bibliographic records downloaded with script characters. The study results revealed that some libraries experience problems with bibliographic utilities when downloading records into their local systems, as for 33.3% of Arabic/Hebrew catalogers.
- The system must support Unicode fully in the bibliographic (and authority) records, on both the client and server levels. Unicode should be an integral part of the system's design. The system should store data natively in Unicode.
- The system must facilitate keyboard or graphical input methods (e.g., virtual keyboards) for inputting and editing JACKPHY* Characters, plus Cyrillic characters. The automated systems' capability to enter or type in vernacular characters was below 50% in each of the scripts of interest to this study.
- The system must provide culturally accurate sorting sequences of non-Roman characters.

[*] JACKPHY includes Chinese, Japanese, Korean, Persian [Farsi], Arabic, Hebrew and Yiddish.

- Localization: The system must involve Unicode support in all areas other than bibliographic access, such as in the user interface, help files, etc. Users should also be allowed to change the interface language on-the-fly. In other words, users should be able to set the prompts in any language on any screen.
- The system should support spell checking and correction of JACKPHY characters.
- The system must also support bi-directional scripts (such as Hebrew and Arabic), including display, character input and editing, and sorting and searching of these scripts. Arabic and Hebrew are unique scripts compared to other non-Roman scripts, as they are written from right to left. The study results revealed very low support for these two scripts in terms of interface language, typing, and search/display, although 76.5% of the respondents indicated that their libraries have future plans concerning the support of these scripts.

5.4 Implications

The researcher attempted, during the course of this study, to determine the current status of support of non-Roman scripts in library OPACs along with academic libraries' future plans toward providing access to original (vernacular) characters in bibliographic records for non-Roman script materials. The collected data has been analyzed and synthesized in a set of recommendations to facilitate decision-making by the following important stakeholders:

1. *Academic and research libraries with sizeable non-Roman script materials*. These libraries should benefit from the results revealed in the study in terms of learning how similar libraries deal with these materials and learning what are the general trends in the near future with regard to bibliographic control and access of the original (vernacular)

characters. Academic libraries need to rethink the way they provide reference services to users interested in non-Roman script materials, especially providing access to transliteration tables, and developing and organizing bibliographic instruction about searching for these materials, either in a romanized or vernacular characters environment.

2. *Bibliographic utilities*. Providing bibliographic records in the vernacular for non-Roman script materials becomes necessary in the major bibliographic utilities in North America, mainly OCLC and RLIN, as they constitute the cooperation ground for different types of libraries. These utilities should gain advantage from the study results with regard to characteristics of these materials in academic and research libraries and how these libraries plan to provide access to these materials in the near future. Of particular interest to these utilities is the exploration of problems encountered by research libraries when obtaining records from the utilities. Developing solutions to assist libraries to better use vernacular records should be further emphasized.
3. *Library automation vendors*. The library automation industry plays a vital role in helping libraries provide effective access to their valuable research collections. It was the intention of this study to propose a set of functional requirements to designers and developers of automated library systems with regard to support of non-Roman scripts in library OPACs. These requirements have been presented above (see 5.3 Functional Requirements for Non-Roman Scripts in Library OPACs). System and interface design implications include recommendations to design multilingual interfaces and to allow searching and displaying vernacular characters. Even automation vendors whose systems have substantial multilingual support should assist their customers with troubleshooting the technical problems related to searching and displaying non-Roman scripts.

4. *Library and Information Science schools.* LIS education needs to pay more systematic attention to issues related to multilingual user populations. Students should be prepared with the appropriate skills to handle functions such as cataloging non-Roman script materials and providing better public services to multilingual/multiscript populations. Managing multilingual support in automated library systems should be included in courses related to library automation, such as computer technology applications, system analysis and design, and human-computer interaction. It was for these reasons that the course "foreign language barriers in information transfer" was organized and introduced in the Department of Library Science of St. John's University during the 1972-73 academic year (Anderson, 1974).

5.5 Recommendations for Future Research

We turn now to the consideration of future research of bibliographic access to non-Roman scripts in library OPACs. Below are some strategies for moving forward. More research is needed on illustrating the significance that academic libraries can have providing bibliographic access to vernacular characters in their OPACs, and assessing how such access could increase the use of the valuable research collections housed in these libraries. In particular, the following themes for future research in vernacular characters in library OPACs need to be emphasized:

- The future of Romanization,
- End users' familiarity with transliteration schemes,
- Re-cataloging, and
- Multilingual authority files.

The future of Romanization. An obvious need for future research is to examine whether Romanization, as a bibliographic tool for non-Roman script materials, should continue to perform the same function in light of developments in bibliographic utilities and technological advances in processing vernacular characters. Many scholars have spoken against Romanization (Weinberg, 1974; Spalding, 1977; and Bachman, 1989). Of special interest in this direction is how librarians see the importance of keeping Romanized bibliographic data along with script characters. Furthermore, the prevalent use of Romanization for Cyrillic cataloging should urgently draw attention as to why Romanization of this script is so acceptable among catalogers. Moreover, a number of catalogers of non-Roman script materials who responded to this study indicated that romanized data will continue to be needed even though users have access to original (vernacular) data. A future, possibly experimental, research may validate this claim.

The specific research questions that arise based on the above analysis are:

- Should bibliographic access to non-Roman scripts in library OPACs continue to be provided in Romanization in light of the availability of vernacular cataloging in bibliographic utilities and the growing Unicode-compliant automated library systems?
- To what extent do catalogers and public service librarians consider the importance of keeping Romanized bibliographic data along with script characters in bibliographic records for non-Roman script materials?
- Is Romanization of Cyrillic easier or more straightforward than Romanization of other non-Roman scripts? If so, is that because of the nature of this script, or the efficiency of ALA/LC Romanization table, or for some other reason?

- Do library users and librarians need access to romanized bibliographic data in addition to original (vernacular) characters? Or will users and librarians depend only on vernacular characters, if available, for bibliographic access to non-Roman script materials?

End users' familiarity with transliteration schemes. Libraries in the Western world have utilized transliteration schemes as the dominant tools for describing bibliographic data for non-Roman script materials. Few studies have been conducted to determine how much library users are familiar with the transliteration schemes used in their libraries. There is much diversity in the transliteration schemes used, and without knowing which scheme is used and also knowing how to use it efficiently and effectively, it is very hard for end users to find the information they need in a romanized catalog. Future research in this direction calls for the inclusion of focus group interviews in data collection and experimental comparisons of romanized-based and vernacular-based bibliographic access environments, especially important for Cyrillic.

The specific research questions that arise based on the above analysis are:

- To what extent are library users aware of and familiar with transliteration schemes used in their libraries? What is the impact of the level of awareness / familiarity and the effectiveness of retrieval of bibliographic records of non-Roman script materials?
- What are the successes and failures of library users when searching for non-Roman script materials when comparing romanized to vernacular bibliographic records?

Re-cataloging. The daunting task of re-cataloging all the romanized bibliographic records to include original scripts has been echoed by a significant number of the librarians responding to this survey. Consequently, there is an urgent need to explore issues related to the retrospective conversion of currently available romanized records into bibliographic records that contain both romanized and vernacular data. Of special interest here is the feasibility of automatic reversibility by extracting original (vernacular) characters based on the romanized bibliographic data.

Another issue is Eilts' (1996) idea of the international exchange of bibliographic data by requiring the records from countries of origin of the materials, e.g., from the book suppliers or the national and research libraries in the various countries.

Some specific research questions that arise based on the above analysis are:

- What are the options / alternatives available to libraries for re-cataloging romanized records to include vernacular scripts?
- What is the effectiveness of retrospective conversion of current romanized records in terms of factors like size of the collection, user population, and expected use?
- What is the feasibility of automatic reversibility from romanized data to original characters? To what extent would such a process be impacted by the nature and characteristics of non-Roman scripts?

Multilingual Authority Files. Authority control facilitates the job of library catalogs in bringing all the works of an author (being a personal or a corporate author) closer to each other. Likewise, it facilitates finding all the works that discuss a certain topic or subject. The intricacies (e.g., benefits, problems, possibilities, etc.) of an OPAC that uses multilingual authority files, which allow one to search in more than one language, are not widely examined. Another

interesting research direction could be exploration (possibilities, problems, etc.) of providing interoperability across multiple authority files, to link and provide switching for displays of authorized headings on an international scale, with the objective of providing the means to access library catalogs on a multilingual basis.

The specific research questions that arise based on the above analysis are:

- What are benefits that users can gain from an OPAC that uses multilingual authority files? What are the design considerations here?
- What is the possibility of linking multiple authority files on an international scale? How will such linking enhance the effective inquiry of library OPACs?

Epilogue...

The study revealed the current practices and future plans of handling non-Roman script materials in library OPACs in academic libraries in the United States. Its findings may suggest new directions in providing better bibliographic access to non-Roman language materials in library OPACs. The report of these findings could also be used as a basis for establishing functional requirements for multilingual support in library OPACs.

The researcher assumed that libraries started cataloging script materials when they implemented automated library systems with the appropriate capabilities. But contrary to expectation, many libraries actually began cataloging these materials in vernacular characters in bibliographic utilities long before their automated systems were capable of handling scripts.

It is hoped that this study of bibliographic access to non-Roman materials gathers interest nationally to an issue that was previously relegated to low priority. The scripts identified in this survey, as indicated by many librarians, perhaps represent the national "consensus" of first-priority scripts but should not be deemed the only scripts needing attention and system support.

APPENDICES

Appendix (A)

The Questionnaire

Dear Library Director

I am a doctoral student in the School of Information Sciences at the University of Pittsburgh, Pennsylvania. I plan to pursue research for my dissertation on the topic of "Non-Roman Scripts in Library Online Public Access Catalogs (OPACs)."

Enclosed is a copy of my questionnaire, divided into five sub-questionnaires (CJK, Middle East, Cyrillic scripts, public services department, and the library OPAC system department). I hope to gather information on current practices of using non-Roman scripts, if any, in your online public access catalog (OPAC), when handling materials in non-Roman scripts (e.g., Arabic, Chinese, Japanese, Korean, Hebrew, Cyrillic, etc.).

I would appreciate your taking the time to distribute the enclosed questionnaires and collect and return them to me in the enclosed postage-paid envelope. Please, refer to each sub-questionnaire's instructions for "*who may complete these questionnaires*." Only a limited number of ARL libraries receive this survey, so your feedback is very important. Please return the completed questionnaire by Sep. 20th 2002.

You may be assured of complete confidentiality. The questionnaire has an identification number for mailing purposes only. This is so that I may check the library name off the mailing list when your questionnaire is returned.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Thank you in advance for your assistance in making this research a success.

Best regards,

Ali Shaker
Department of Library and Information Science
School of Information Sciences
University of Pittsburgh, PA (USA)

Enclosed is a copy of my questionnaire, divided into five sub-questionnaires

Questionnaire on CJK Scripts in Library OPACs

General Instructions

What is this?

This brief questionnaire is an important part of a survey that is being conducted to selected ARL academic libraries. The purpose of the study is to identify the bibliographic access to non-Roman script materials in library OPACs along with the services provided to facilitate access to these materials.

The questionnaire, containing 24 questions, is divided into three parts:

Part I. The Collection Profile

Focused on characteristics of Chinese, Japanese, and Korean (CJK) script materials, and is designed to identify practices in organizing, and shelving these collections.

Part II. The OPAC Profile

Intended to assess the availability of CJK scripts' support features in library OPACs: (1) the capability to change the interface language according to user preferences, (2) the capability to search for/display these scripts in the vernacular

Part III. The Staff Profile

Aimed to gather personnel data about librarians working with CJK script materials. Data is requested about catalogers in order to identify the availability of librarians with language skills and also to determine whether they experience any problems with Romanization of CJK scripts.

Sizeable CJK collections are defined based on <u>at least one</u> of the following three criteria:

1. The collection has a specific identity (e.g., East Asian Collection, Chinese collection, Japanese collection, Korean collection), whether it is separated from the entire collection or integrated within;
2. The collection development policy states that "area studies" is a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and
3. The library has librarians whose main responsibilities are to acquire/catalog CJK script materials.

Who may complete this questionnaire?

The head of East Asian collection in case of the three scripts represent one collection. In case there is one or more separate collections of each script, then the librarian in charge of the most sizeable collection may complete this questionnaire.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Part I

CJK Script Materials: The Collection Profile

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

1. Our library has sizeable collection(s)* of CJK script materials, as follows: (Please circle all that apply)

   a. Chinese b. Japanese c. Korean d. None

If "None" is the answer to the first question, please stop here.
Thanks for your time.
Please see return information on the last page.

Otherwise, please continue to the next question.

2. In terms of integration with/separation from the Roman script collection in our library: (Please skip any collection if it is not applicable)

   a. The Chinese Collection is:……. Integrated Separated
   b. The Japanese Collection is:…… Integrated Separated
   c. The Korean Collection is:……... Integrated Separated

3. Source of copy catalog records for non-Roman script materials:
(Please check all that apply)

   a. Chinese……. OCLC RLIN OTHER
   b. Japanese…… OCLC RLIN OTHER
   c. Korean…..…… OCLC RLIN OTHER

If "OTHER" source is used than OCLC and/or RLIN, please give details below:

---

* Please refer to "General Instructions" page for a definition of "sizeable collection" as it is used in this study.

4. The library stores the bibliographic records of these materials:

Key: Please write "A" if → in the main library OPAC
"B" if → in separate databases/files
"C" if → in paper file (e.g., card catalog, book-like catalog, etc.)

(Please skip any collection if it is not applicable)

a. For the Chinese Collection [ ]
b. For the Japanese Collection [ ]
c. For the Korean Collection [ ]

5. Our library obtains (e.g., copy cataloging) bibliographic records for CJK script materials presented in which of the following ways:

Key: Please write "A" if → romanized bibliographic data only
"B" if → combination of romanized and vernacular data

(Please skip any collection if it is not applicable)

a. For the Chinese Collection [ ]
b. For the Japanese Collection [ ]
c. For the Korean Collection [ ]

6. Our library creates (e.g., original cataloging) bibliographic records for CJK script materials in which of the following ways:

Key: Please write "A" if → Romanization only
"B" if → combination of romanized and vernacular data

(Please skip any collection if it is not applicable)

a. For the Chinese Collection [ ]
b. For the Japanese Collection [ ]
c. For the Korean Collection [ ]

If Romanization is the only method of cataloging all or some of the CJK script materials, please answer the following two questions:

7. Please specify which scheme(s) (e.g., Wade-Giles, Pinyin) is (are) currently used:
Answer with "Not applicable" where original scripts are used.
Where scheme for certain script(s) has changed over the years, please give details.

a. For the Chinese Collection ----------
b. For the Japanese Collection----------
c. For the Korean Collection ----------

8. What are the main reasons for cataloging CJK materials using Romanization? (Please circle all that apply)

   a. Romanization is adequate for library goals
   b. No staff with language skills to catalog in the vernacular
   c. The current automated system does not support any other way (e.g., vernacular characters representation)
   d. The high cost associated with original (vernacular) cataloging
   e. Problems with obtaining the records with scripts from the bibliographic utilities (e.g., OCLC, RLIN)
   f. Others (please specify) ----------

If original scripts are used for cataloging all or some of the CJK script materials, please answer the following two questions:

9. When did you start cataloging CJK script materials in the vernacular in your OPAC?

   a. For Chinese script: ----------
   b. For Japanese script: ----------
   c. For Korean script: ----------

10. What was the reason(s) behind using CJK vernacular characters? (Circle all that apply)

    a. Providing better access for users than romanized characters
    b. Already having or recruiting new librarians with language skills
    c. Implementing an automated system that makes this possible
    d. The vernacular bibliographic records became available from bibliographic utilities (OCLC, RLIN, etc.)
    e. Other(s) please specify: ----------

Part II

Bibliographic Access to CJK Script Materials: The OPAC Profile

<u>Directions:</u> please answer each question by circling the number of the response that matches your answer best. Give only <u>one</u> response, unless otherwise instructed.

11. The system, as it is used in the library, permits choosing the interface language* based on the user's (librarians, end users) preferences:

    a. Chinese interface……...□ Yes □ No
    b. Japanese interface..……...□ Yes □ No
    c. Korean interface…….……..…□ Yes □ No

12. The system, as it is used in the library, permits entering bibliographic data of CJK script materials in their vernacular characters (the cataloging module).

    a. Chinese.……...□ Yes □ No
    b. Japanese..……...□ Yes □ No
    c. Korean…….……..…□ Yes □ No

13. The system, as it is used in the library, permits searching and displaying bibliographic data of CJK script materials in their vernacular characters (by the end user)
    Key: write "**1**"→ if (searching and display)
    "**2**"→ if (display only)
    "**3**"→ if (this is not possible)

    a. Chinese [ ] b. Japanese [ ] c. Korean [ ]

14. Which of the following fields of the bibliographic record is romanized, and which is in the vernacular script:
- Please, write "**<u>R</u>**" if the field is romanized,
    "**<u>V</u>**" if field is in the vernacular,
    "**<u>B</u>**" if the field is both romanized and in the vernacular.

    [ ] Title
    [ ] Statement of Responsibility
    [ ] Publication (place: publisher, year)
    [ ] Series
    [ ] Subject Headings

* The interface language of the automated library system allows the user to choose the language of menus, dialog boxes, help screens, etc., that is desired. However, this does not mean changing the language of the keyboard. For example, a user can choose Russian as the language interface while still working in English.

If the current OPAC contains only romanized bibliographic data for CJK script materials, please answer questions # 15 to # 17, then skip to question # 23.

15. To what extent do you agree with the following statement, "*Libraries should catalog all CJK materials in the original language and script*"?

□ Strongly agree □ Agree □ Disagree □ Strongly disagree

Please explain why you chose your selection above:

---

16. Will the library consider original (vernacular) characters in the CJK bibliographic records in the near future (within 5 years)?

a. Yes b. No

Comments: ---

---

17. What are the current obstacles to using the original (vernacular) characters in your library OPACs, in the case that they are not used for any of the scripts? (Please circle all that apply)

a. No barrier, we are just fine with Romanization
b. The current staff lacks necessary language skills
c. It may be costly
d. The current OPAC system does not support vernacular characters
e. Difficulties with obtaining the records with vernacular characters from the bibliographic utilities (e.g., OCLC, RLIN)
f. Other(s) please specify ---

If the current OPAC contains bibliographic records that contain a combination of romanized and original script bibliographic data for CJK script materials, please answer questions # 18 to # 22, then continue with question # 23.

18. If your OPAC contains original (vernacular) scripts, the library provides access to these bibliographic records:

a. In all library public computers
b. In special computers dedicated for these scripts→ Skip to question # 20

19. If "A" was your answer to the previous question, please explain in more detail the current settings (e.g., how are the computers equipped to handle original scripts)

---

20. If "B" was your answer to question # 18, please explain in more detail the current settings (e.g., how are these special computers equipped to handle original scripts)

---

21. If original (vernacular) characters are available in your library OPAC, please select the level of availability:

Key: <u>All</u> → vernacular characters available in all bibliographic records for this script
<u>Some</u> → vernacular characters available in some of the bibliographic records

a. Chinese…… □ All □ Some
b. Japanese……□ All □ Some
c. Korean……..□ All □ Some

22. If original (vernacular) characters are available in <u>some of the bibliographic records</u>, does the library plan to convert the other non-vernacular into vernacular records in the near future (within 5 years)?

a. Chinese…… □ Yes □ No
b. Japanese……□ Yes □ No
c. Korean……..□ Yes □ No

Comments: **----------**

Part III

Personnel Handling CJK Script Materials: The Staff Profile

The purpose of this part is to obtain some background information that will be used in the statistical analysis of this study. Please circle the item(s) that apply.

23. Our library has catalogers, with language skills, specifically for the following CJK script(s) materials: (Please circle all that apply)

    a. Chinese script
    b. Japanese script
    c. Korean script

24. What are the problems, if any, encountered by library catalogers with regard to using Romanization for CJK materials?

→ Skip if Romanization is not used for cataloging CJK script materials.

    a. Multiplicity of schemes for certain script(s)
    b. Difficulty in applying/using the scheme(s)
    c. The long time it takes to catalog a book
    d. Other(s) Please specify ----------

    ----
    e. None

Thank you for your help!

Please return the questionnaire to the library director.

Have a great day!

Questionnaire on Middle East Scripts in Library OPACs

General Instructions

What is this?

This brief questionnaire is an important part of a survey that is being conducted to selected ARL academic libraries. The purpose of the study is to identify the bibliographic access to non-Roman script materials in library OPACs along with the services provided to facilitate access to these materials.

The questionnaire, containing 24 questions, is divided into three parts:

Part I. The Collection Profile

Focused on characteristics of Middle East script materials, and is designed to identify practices in organizing, and shelving these collections.

Part II. The OPAC Profile

Intended to assess the availability of Arabic/Hebrew scripts' support features in library OPACs: (1) the capability to change the interface language (to Arabic/Hebrew) according to user preferences, (2) the capability to search for/display these scripts in the vernacular.

Part III. The Staff Profile

Aimed to gather personnel data about librarians working with Middle East script materials. Data is requested about catalogers in order to identify the availability of librarians with language skills and also to determine whether they experience any problems with Romanization of these scripts.

Middle East scripts of interest to this questionnaire are: Arabic and Hebrew

Sizeable Middle East collections are defined based on at least one of the following three criteria:

1. The collection has a specific identity (e.g., Middle East collection, Islamic studies collection, Hebraica collection), whether it is separated from the entire collection or integrated within;
2. The collection development policy states that "area studies" is a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and
3. The library has librarians whose main responsibilities are to acquire/catalog Middle East script materials.

Who may complete this questionnaire?

The head of Middle East collection in case of the two scripts represent one collection. In case there is one or more separate collections of each script, then the librarian in charge for the most sizeable collection may complete this questionnaire.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Part I

Middle East Script Materials: The Collection Profile

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

1. Our library has sizeable collection(s)* of Arabic/Hebrew script materials, as follows: (Please circle all that apply)

   a. Arabic b. Hebrew c. None

If "None" is the answer to the first question, please stop here.
Thanks for your time.
Please see return information on the last page.

Otherwise, please continue to the next question.

2. In terms of integration with/separation from the Roman script collection in our library: (Please skip any collection if it is not applicable)

   a. The Arabic Collection is:……. □ Integrated □ Separated
   b. The Hebrew Collection is:…… □ Integrated □ Separated

3. Source of copy catalog records for these script materials:
(Please check all that apply)

   a. Arabic……..…….. □ OCLC □ RLIN □ OTHER
   b. Hebrew……..…… □ OCLC □ RLIN □ OTHER

If "OTHER" source is used than OCLC or RLIC, please give details below:

---

4. The library stores the bibliographic records of these materials:
Key: Please write "A" if → in the main library OPAC
"B" if → in separate databases/files
"C" if → in paper file (e.g., card catalog, book-like catalog, etc.)
(Please skip any collection if it is not applicable)

   a. For the Arabic Collection [ ]
   b. For the Hebrew Collection [ ]

* Please refer to "General Instructions" page for a definition of "sizeable collection" as it is used in this study.

5. Our library obtains (e.g., copy cataloging) bibliographic records for Arabic/Hebrew scripts materials presented in which of the following ways:

Key: Please write "A" if → romanized bibliographic data only
"B" if → combination of romanized and vernacular data
(Please skip any collection if it is not applicable)

a. For the Arabic Collection [ ]
b. For the Hebrew Collection [ ]

6. Our library creates (e.g., original cataloging) bibliographic records for Arabic/Hebrew script materials in which of the following ways:

Key: Please write "A" if → romanized bibliographic data only
"B" if → combination of romanized and vernacular data
(Please skip any collection if it is not applicable)

a. For the Arabic Collection [ ]
b. For the Hebrew Collection [ ]

If Romanization is the only method of cataloging all or some of the Arabic/Hebrew script materials, please answer the following two questions:

7. Please specify which scheme(s) (e.g., ALA/LC Romanization tables) is (are) currently used:
Answer with "Not applicable" where original scripts are used.
Where scheme for certain script(s) has changed over the years, please give details.

a. For the Arabic Collection ----------
b. For the Hebrew Collection----------

8. What are the main reasons for cataloging Arabic/Hebrew materials using Romanization? (Please circle all that apply)

a. Romanization is adequate for library goals
b. No staff with language skills to catalog in the vernacular
c. The current automated system does not support any other way (e.g., vernacular characters representation)
d. The high cost associated with original (vernacular) cataloging
e. Problems with obtaining the records with scripts from the bibliographic utilities (e.g., OCLC, RLIN)
f. Others (please specify) ----------

If original scripts are used for cataloging all or some of the Arabic/Hebrew script materials, please answer the following two questions:

9. When did you start cataloging Arabic/Hebrew script materials in the vernacular in your OPAC?

    a. For Arabic script: ----------
    b. For Hebrew script: ----------

10. What was the reason(s) behind using vernacular characters? (Circle all that apply)

    a. Providing better access for users than romanized characters
    b. Already having or recruiting new librarians with language skills
    c. Implementing an automated system that makes this possible
    d. The vernacular bibliographic records became available from bibliographic utilities (OCLC, RLIN, etc.)
    e. Other(s) please specify: ----------

Part II

Bibliographic Access to Middle East Script Materials: The OPAC Profile

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

11. The system, as it is used in the library, permits choosing the interface language* based on the user's (librarians, end users) preferences:

    a. Arabic interface……□ Yes □ No
    b. Hebrew interface……□ Yes □ No

12. The system, as it is used in the library, permits entering bibliographic data of Arabic/Hebrew script materials in their vernacular characters (the cataloging module).

    a. Arabic…….....….....□ Yes □ No
    b. Hebrew…….....…□ Yes □ No

13. The system, as it is used in the library, permits searching and displaying bibliographic data of Arabic/Hebrew script materials in their vernacular characters (by the end user)
    Key: write "**1**"→ if (searching and display)
    "**2**"→ if (display only)
    "**3**" → if (this is not possible)

    a. Arabic Script [ ]
    b. Hebrew Script [ ]

14. Which of the following fields of the bibliographic record is romanized, and which is in the vernacular script:

- Please, write "**R**" if the field is romanized,
  "**V**" if the field is in the vernacular,
  "**B**" if the field is both romanized and in the vernacular.

    [ ] Title
    [ ] Statement of Responsibility
    [ ] Publication (place: publisher, year)
    [ ] Series
    [ ] Subject Headings

* The interface language of the automated library system allows the user to choose the language of menus, dialog boxes, help screens, etc., that is desired. However, this does not mean changing the language of the keyboard. For example, a user can choose Russian as the language interface while still working in English.

If the current OPAC contains only romanized bibliographic data for non-Roman script materials, please answer questions # 15 to # 17, then skip to question # 23.

15. To what extent do you agree with the following statement, "*Libraries should catalog all Arabic/Hebrew materials in the original language and script*"?

□ Strongly agree □ Agree □ Disagree □ Strongly disagree

Please explain why you chose your selection above:

---

16. Will the library consider original (vernacular) characters in the bibliographic records in the near future (within 5 years)?

a. Yes b. No

Comments: ---

---

17. What are the current obstacles to using the original (vernacular) characters in your library OPACs, in the case that they are not used for any of the scripts? (Please circle all that apply)

a. No barrier, we are just fine with Romanization
b. The current staff lacks necessary language skills
c. It may be costly
d. The current OPAC system does not support vernacular characters
e. Difficulties with obtaining the records with vernacular characters from the bibliographic utilities (e.g., OCLC, RLIN)
f. Other(s) please specify ---

If the current OPAC contains bibliographic records that contain a combination of romanized and original script bibliographic data for Arabic/Hebrew script materials, please answer questions # 18 to # 22, then continue with question # 23.

18. If your OPAC contains original (vernacular) scripts, the library provides access to these bibliographic records:

a. In all library public computers
b. In special computers dedicated for these scripts→ Skip to question # 20

19. If "A" was your answer to the previous question, please explain in more detail the current settings (e.g., how are the computers equipped to handle original scripts)

---

20. If "B" was your answer to question # 18, please explain in more detail the current settings (e.g., how are these special computers equipped to handle original scripts)

---

21. If original (vernacular) characters are available in your library OPAC, please select the level of availability:

Key: <u>All</u> → vernacular characters available in all bibliographic records for this script
<u>Some</u> → vernacular characters available in some of the bibliographic records

a. Arabic……□ All □ Some
b. Hebrew……..□ All □ Some

22. If original (vernacular) characters are available in <u>some of the bibliographic records</u>, does the library plan to convert the other non-vernacular into vernacular records in the near future (within 5 years)?

a. Arabic……□ Yes □ No
b. Hebrew……..□ Yes □ No

Comments: ----------

Part III

Personnel Handling Non-Roman Script Materials: The Staff Profile

The purpose of this part is to obtain some background information that will be used in the statistical analysis of this study. Please circle the item(s) that apply.

23. Our library has catalogers, with language skills, specifically for the following Arabic/Hebrew script(s) materials: (Please circle all that apply)

    a. Arabic script
    b. Hebrew script

24. What are the problems, if any, encountered by library catalogers with regard to using Romanization for Arabic/Hebrew materials?

→ Skip if Romanization is not used for cataloging non-Roman script materials.

    a. Multiplicity of schemes for certain script(s)
    b. Difficulty in applying/using the scheme(s)
    c. The long time it takes to catalog a book
    d. Other(s) Please specify ----------

    ----
    e. None

Thank you for your help!

Please return the questionnaire to the library director

Have a great day!

Questionnaire of Cyrillic Scripts in Library OPACs

General Instructions

What is this?

This brief questionnaire is an important part of a survey that is being conducted to selected ARL academic libraries. The purpose of the study is to identify the bibliographic access to non-Roman script materials in library OPACs along with the services provided to facilitate access to these materials.

The questionnaire, contains 24 questions, is divided into three parts:

Part I. The Collection Profile

Focused on characteristics of Cyrillic script materials, and is designed to identify practices in organizing, and shelving these collections.

Part II. The OPAC Profile

Intended to assess the availability of Cyrillic scripts' support features in library OPACs: (1) the capability to change the interface language according to user preferences, (2) the capability to search for/display these scripts in the vernacular

Part III. The Staff Profile

Aimed to gather personnel data about librarians working with Cyrillic script materials. Data is requested about catalogers in order to identify the availability of librarians with language skills and also to determine whether they experience any problems with Romanization of these scripts.

Cyrillic scripts of interest to this questionnaire are: Russian

Sizeable Cyrillic collections are defined based on at least one of the following three criteria:

1. The collection has a specific identity (e.g., Slavic and East European collection, Russian studies), whether it is separated from the entire collection or integrated within;
2. The collection development policy states that "area studies" is a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and
3. The library has librarians whose main responsibilities are to acquire/catalog Cyrillic script materials.

Who may complete this questionnaire?

The head of Slavic studies collection or the senior librarian may complete this questionnaire.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Part I

Cyrillic Script Materials: The Collection Profile

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

1. Our library has sizeable Cyrillic (e.g., Russian) collection*.

a. Yes b. No

If "No" is the answer to the first question, please stop here.
Thanks for your time.
Please see return information on the last page.

Otherwise, please continue to the next question.

2. In terms of integration with/separation from the Roman script collection in our library, the Cyrillic collection is:

a. Integrated b. Separated

3. Source of copy catalog records for the Cyrillic script materials:
(Please check all that apply)

a. OCLC b. RLIN c. OTHER

If "OTHER" source is used than OCLC and/or RLIC, please give details below:

---

4. The library stores the bibliographic records of these materials:

a. in the main library OPAC
b. in separate databases/files
c. in paper file (e.g., card catalog, book-like catalog, etc.)

5. Our library obtains (e.g., copy cataloging) bibliographic records for Cyrillic materials presented in which of the following ways:

a. Romanization only
b. Combination of romanized and vernacular data

* Please refer to "General Instructions" page for a definition of "sizeable collection" as it is used in this study.

6. Our library creates (e.g., original cataloging) bibliographic records for Cyrillic materials in which of the following ways:

    a. Romanization only
    b. Combination of romanized and vernacular data

If Romanization is the only method of cataloging all or some of the Cyrillic materials, please answer the following two questions:

7. Please specify which scheme(s) (e.g., ALA/LC Romanization tables) is (are) currently used for romanizing the Cyrillic materials:
   Where the scheme has changed over the years, please give details.

   Scheme name: ----------
   ----

8. What are the main reasons for cataloging Cyrillic materials using Romanization? (Please circle all that apply)

    a. Romanization is adequate for library goals
    b. No staff with language skills to catalog in the vernacular
    c. The current automated system does not support any other way (e.g., vernacular characters representation)
    d. The high cost associated with original (vernacular) cataloging
    e. Problems with obtaining the records with scripts from the bibliographic utilities (e.g., OCLC, RLIN)
    f. Others (please specify) ----------

If original scripts are used for cataloging all or some of the Cyrillic materials, please answer the following two questions:

9. When did you start cataloging Cyrillic script materials in the vernacular in your OPAC?
   Please, enter year here: ----------

10. What was the reason(s) behind using vernacular characters? (Circle all that apply)

    a. Providing better access for users than romanized characters
    b. Already having or recruiting new librarians with language skills
    c. Implementing an automated system that makes this possible
    d. The vernacular bibliographic records became available from bibliographic utilities (OCLC, RLIN, etc.)
    e. Other(s) please specify: ----------

Part II

Bibliographic Access to Cyrillic Script Materials: The OPAC Profile

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

11. The system, as it is used in the library, permits choosing a Cyrillic interface language* based on the user's (librarians, end users) preferences:

a. Yes b. No

12. The system, as it is used in the library, permits entering bibliographic data of Cyrillic script materials in their vernacular characters (the cataloging module).

a. Yes b. No

13. The system, as it is used in the library, permits searching and displaying bibliographic data of Cyrillic script materials in their vernacular characters (by the end user)

a. searching and display b. display only c. this is not possible

14. Which of the following fields of the bibliographic record is romanized, and which is in the vernacular script:

- Please, write "**R**" if the field is romanized,
  "**V**" if the field is in the vernacular,
  "**B**" if the field is both romanized and in the vernacular.

[ ] Title
[ ] Statement of Responsibility
[ ] Publication (place: publisher, year)
[ ] Series
[ ] Subject Headings

* The interface language of the automated library system allows the user to choose the language of menus, dialog boxes, help screens, etc., that is desired. However, this does not mean changing the language of the keyboard. For example, a user can choose Russian as the language interface while still working in English.

If the current OPAC contains only romanized bibliographic data for non-Roman script materials, please answer questions # 15 to # 17, then skip to question # 23.

15. To what extent do you agree with the following statement, "*Libraries should catalog all Cyrillic materials in the original language and script*"?

□ Strongly agree □ Agree □ Disagree □ Strongly disagree

Please explain why you chose your selection above:

---

16. Will the library consider original (vernacular) characters in the bibliographic records in the near future (within 5 years)?

a. Yes b. No

Comments: ---

---

17. What are the current obstacles to using the original (vernacular) Cyrillic characters in your library OPACs? (Please circle all that apply)

a. No barrier, we are just fine with Romanization
b. The current staff lacks necessary language skills
c. It may be costly
d. The current OPAC system does not support vernacular characters
e. Difficulties with obtaining the records with vernacular characters from the bibliographic utilities (e.g., OCLC, RLIN)
f. Other(s) please specify ---

If the current OPAC contains bibliographic records that contain a combination of romanized and original script bibliographic data for Cyrillic script materials, please answer questions # 18 to # 22, then continue with question # 23.

18. If your OPAC contains original (vernacular) scripts, the library provides access to these bibliographic records:

a. In all library public computers
b. In special computers dedicated for these scripts→ Skip to question # 20

19. If "A" was your answer to the previous question, please explain in more detail the current settings (e.g., how are the computers equipped to handle original scripts)

---

20. If "B" was your answer to question # 18, please explain in more detail the current settings (e.g., how are these special computers equipped to handle original scripts)

---

21. If original (vernacular) characters are available in your library OPAC, please select the level of availability:

Key: <u>All</u> → vernacular characters available in all bibliographic records for this script
<u>Some</u> → vernacular characters available in some of the bibliographic records

a. Cyrillic……..□ All □ Some

22. If original (vernacular) characters are available in <u>some of the bibliographic records</u>, does the library plan to convert the other non-vernacular into vernacular records in the near future (within 5 years)?

a. Yes b. No

Comments: **----------**

Part III

Personnel Handling Cyrillic Script Materials: The Staff Profile

The purpose of this part is to obtain some background information that will be used in the statistical analysis of this study. Please circle the item(s) that apply.

23. Our library has catalogers, with language skills, specifically for the Cyrillic script materials:

    a. Yes b. No

24. What are the problems, if any, encountered by library catalogers with regard to using Romanization for Cyrillic materials?

→ Skip if Romanization is not used for cataloging non-Roman script materials.

a. Multiplicity of schemes for certain script(s)
b. Difficulty in applying/using the scheme(s)
c. The long time it takes to catalog a book
d. Other(s) Please specify ----------

   ----
e. None

Thank you for your help!

Please return the questionnaire to the library director

Have a great day!

Non-Roman Scripts in Library OPACs: The Public Services

General Instructions

What is this?

This brief questionnaire is an important part of a survey that is being conducted to selected ARL academic libraries. The purpose of the study is to identify the bibliographic access to non-Roman script materials in library OPACs along with the services provided to facilitate access to these materials.

Non-Roman scripts of interest to this questionnaire are:
Arabic, Hebrew, CJK, and Russian.

Sizeable non-Roman script collections are defined based on at least one of the following three criteria:

1. The collection has a specific identity (e.g., East Asian Collection, Middle East collection, Slavic and East European collection), whether it is separated from the entire collection or integrated within;

a. The collection development policy states that "area studies" is a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and

2. The library has librarians whose main responsibilities are to acquire/catalog non-Roman script materials.

Who may complete this questionnaire?
The head of public services department or the senior reference librarian may complete this questionnaire.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

1. Do public service librarians (e.g., reference librarians) and end users have access to the transliteration table(s) (e.g., ALA/LC Romanization Tables)? (Select only one answer)

a. Librarians only b. End users only c. Both d. None

2. The library offers instruction or training to end users on searching for non-Roman materials in the library OPAC, regardless if Romanization or original script is used.

a. Yes b. No

3. Library users have access to instructional materials (e.g., handout, information on the library's web site, on the help file within the OPAC, etc.) for searching for non-Roman materials in the library OPAC

a. Yes b. No

4. Our library has public service librarians (e.g., reference librarians) fluent in the following non-Roman Language(s): (Please circle all that apply)

a. Chinese
b. Japanese
c. Korean
d. Arabic
e. Hebrew
f. Cyrillic
g. None

5. What are the problems, if any, encountered by the reference librarians with regard to Romanization?

→ Skip if Romanization is not used for retrieving all non-Roman script materials.

a. Multiplicity of schemes for certain script(s)
b. Difficulty in applying/using the scheme(s)
c. The time it takes to search for books at a public service point (e.g., the reference desk)
d. Other(s) Please specify ----------
e. None

6. Do you have any more comments you would like to add:

Thank you for your help!
Please return the questionnaire to the library director
Have a great day!

Questionnaire of Non-Roman Scripts in Library OPACs: The Systems Department

General Instructions

What is this?

This brief questionnaire is an important part of a survey that is being conducted to selected ARL academic libraries. The purpose of the study is to identify the bibliographic access to non-Roman script materials in library OPACs along with the services provided to facilitate access to these materials.

Non-Roman scripts of interest to this questionnaire are:
Arabic, Hebrew, CJK, and Russian.

Sizeable non-Roman script collections are defined based on at least one of the following three criteria:

1. The collection has a specific identity (e.g., East Asian Collection, Middle East collection, Slavic and East European collection), whether it is separated from the entire collection or integrated within;

a. The collection development policy states that "area studies" is a focus of the collection to support the teaching and research of the academic programs in different fields of area studies; and

2. The library has librarians whose main responsibilities are to acquire/catalog non-Roman script materials.

Who may complete this questionnaire?

The head of systems department or the senior systems librarian may complete this questionnaire.

Should you have any questions regarding the questionnaire, please contact me via e-mail: aks13@pitt.edu

Directions: please answer each question by circling the number of the response that matches your answer best. Give only one response, unless otherwise instructed.

1. Please give the following information about the current automated library system in your library:

    a. Name of the vendor: ----------
    b. Name of the system: ----------
    c. Year system installed: ----------
    d. Current version: ----------
    e. Year current version installed: ----------

2. In which operating system does the current automated library system operate in your library?

Staff devices (e.g., catalogers)	End user devices (e.g., lib. OPAC)
....	....

3. When selecting the current automated library system, we considered supporting non-Roman scripts as important functionality (entering/displaying original 'vernacular' characters):

    a. Yes b. No

4. If "Yes", Please give details:

---

5. If "No", would you consider this support as essential functionality in the near future (within 5 years)?

    a. Yes b. No

6. Do you have any more comments you would like to add:

Thank you for your help!
Please return the questionnaire to the library director
Have a great day!

Appendix (B)

Validation Checklist

To: PILOT STUDY PARTICIPANTS
From: Ali Shaker
16 Melba Pl. # 1
Pittsburgh, PA 15213

After you complete the questionnaire, would you please answer the following questions concerning the format and content of the questionnaire. Thank you for taking time to complete the questionnaire and the checklist. Please return this form with your completed questionnaire.

1. Was the purpose of the research clearly described in the cover letter? □ Yes □ No
2. How long did it take you to complete the questionnaire? _____
3. Was the length of time acceptable? □ Yes □ No
4. Were the directions clear? □ Yes □ No
5. Were there any unclear questions?
   Please indicate the items and explain _____
   _____
   _____
6. Did you find any questions or items objectionable? □ Yes □ No
   Please indicate the items and explain
   _____
   _____
7. Did you question the value of any item(s)? □ Yes □ No
   Please indicate the items and explain
   _____
   _____
8. Please use this space, and the back of this page, to make any other comments or suggestions about the questionnaire.
   _____
   _____

Thanks a lot for your cooperation!

Appendix (C)

Sample Frame

East Asian Collections	University of Arizona Arizona State University Libraries Brigham Young University Brown University University of California - Berkeley University of California - Davis University of California - Irvine University of California - Los Angeles University of California - Riverside University of California - San Diego University of California - Santa Barbara University of Chicago University of Cincinnati Libraries University of Colorado Colorado State University Columbia University Cornell University University of Delaware Duke University Emory University University of Florida Georgetown University University of Georgia Harvard University University of Hawaii University of Illinois - Chicago University of Illinois - Urbana-Champaign Indiana University University of Iowa Johns Hopkins University University of Kansas University of Maryland University of Massachusetts University of Michigan Michigan State University University of Minnesota University of North Carolina - Chapel Hill North Carolina State University Ohio University Ohio State University University of Oregon University of Pennsylvania Pennsylvania State University Libraries University of Pittsburgh Princeton University Purdue University Rutgers University University of Southern California Syracuse University University of Tennessee - Knoxville University of Texas - Austin Tulane University Vanderbilt University University of Virginia Washington State University Washington University - St. Louis Wayne State University University of Wisconsin - Madison Yale University
Middle East Collections	University of Arizona University of California - Berkeley University of California - Riverside University of Chicago Columbia University Cornell University Duke University Emory University Georgetown University George Washington University Harvard University University of Illinois - Chicago Indiana University Johns Hopkins University University of Michigan Michigan State University New York University University of North Carolina - Chapel Hill Ohio State University University of Pennsylvania Pennsylvania State University Libraries Princeton University Stanford University University of Texas - Austin University of Utah University of Virginia University of Washington Washington University - St. Louis Wayne State University Yale University
Judaica and Hebraica Collections	University of California - Berkeley University of Chicago University of Cincinnati Libraries Columbia University University of Connecticut University of Delaware Emory University University of Florida Georgetown University George Washington University University of Illinois - Chicago Northwestern University Library Ohio State University University of Pennsylvania Pennsylvania State University Libraries Princeton University Stanford University State University of New York - Buffalo University of Texas - Austin Tulane University University of Washington Yale University
Slavic Collection	University of Alabama Libraries University of Arizona Arizona State University Libraries Brown University University of California - Berkeley University of California - Riverside University of Chicago University of Cincinnati Libraries Colorado State University Cornell University University of Delaware Duke University Emory University University of Georgia Harvard University University of Hawaii University of Illinois - Chicago University of Illinois - Urbana-Champaign Indiana University University of Kansas University of Michigan University of North Carolina - Chapel Hill Northwestern University Library University of Notre Dame University of Pittsburgh Princeton University Rice University University of Rochester Stanford University Syracuse University University of Texas - Austin Vanderbilt University University of Virginia Virginia Tech University of Washington University of Wisconsin - Madison Yale University

Appendix (D)

Institutional Review Board (IRB) Approval

University of Pittsburgh
Institutional Review Board

Exempt and Expedited Reviews
Christopher M. Ryan, Ph.D., Vice Chair

3500 Fifth Avenue
Suite 105
Pittsburgh, PA 15213
Phone: 412.578.3424
Fax: 412.578.8566
e-mail: irbexempt@msx.upmc.edu

TO: Ali Shaker

FROM: Christopher M. Ryan, Ph.D., Vice Chair

DATE: September 4, 2002

PROTOCOL: Bibliographical Access to Non-Roman Scripts in Library OPACs: A Study of Selected ARL Libraries in the United States

IRB Number: 020849

The above-referenced protocol has been reviewed by the University of Pittsburgh Institutional Review Board. This protocol meets all the necessary requirements and is hereby designated as "exempt" under section 45 CFR 46.101(b)(2). Exempt protocols must be re-reviewed every three years. If you wish to continue the research after that time, a new application must be submitted.

- If any modifications are made to this project, please submit an 'exempt modification' form to the IRB.
- Please advise the IRB when your project has been completed so that it may be officially terminated in the IRB database.
- This research study may be audited by the University of Pittsburgh Research Conduct and Compliance Office.

Approval Date: September 4, 2002
Renewal Date: September 3, 2005

References

Agenbroad, James Edward. (1992) Nonromanization: Prospects for Improving Automated Cataloging of Items in Other Writing Systems. Version of a paper presented at a meeting of the *Library of Congress Cataloging Forum* (Washington, D C., July 22, 1991). (ERIC Document Reproduction Service No. ED354 915)

Aissing, Alena L. (1995) Cyrillic Transliteration and Its Users. *College and Research Libraries*, v. 56, n. 3, 207-19

ALA-LC Romanization tables: transliteration schemes for non- Roman scripts. (1991) Approved by the Library of Congress and the American Library Association; tables compiled and edited by Randall K. Barry. Washington: Cataloging Distribution Service, Library of Congress. 216

Aliprand, Joan M. (1992a) Arabic Script on RLIN. *Library Hi Tech,* vol. 10, no. 4, 59-80

Aliprand, Joan M. (1992b) Nonroman Scripts in the Bibliographic Environment. *Information Technology and Libraries*, vol. 11, no. 2, 105-119

Aliprand, Joan M. (1993) Linkage in USMARC Bibliographic Records. *Cataloging & Classification Quarterly,* vol. 16, no. 1, 5-37

Anderson, Dorothy. (1982) *UBC: A Survey of Universal Bibliographic Control*. London: IFLA International Office for UBC. (Occasional Papers no. 10). 24-5

Anderson, James D. (1974) Foreign Language Barriers in Information Transfer. *Journal of Education for Librarianship*, vol. 14, no. 3, 171-84

Anglo-American Cataloging Rules. (1998) Prepared by the American Library Association, the Library of Congress, the Library Association, and the Canadian Library Association. 2nd. Ed., 1998 Revision. Chicago: American Library Association. Glossary

ARL. *Statement on Qualifications for Membership in the Association of Research Libraries*. Retrieved 6/18/2002 from http://www.arl.org/stats/qualify.html

Arsenault, Clement. (2002) Pinyin Romanization for OPAC Retrieval—is Everyone Being Served? *Information Technology and Libraries,* vol. 21, no. 2, 45-50

Ashoor, Saleh and Abdus Sattar Chaudhry. (1994) Bi-Lingual Bibliographic Software for Processing Library Materials in the Arabic Language. *Program,* vol. 28, no. 2, 167-175

Babu, B. Ramesh and O'Brien, A. (2000) Web OPAC Interfaces: an Overview. *The Electronic Library,* vol. 18, no. 5, 316-330

Bachman Marie Louise. (1989) Translitteration—Ett Bibliografisk Dilemma. Transliteration—A Bibliographic Dilemma. *Tidskrift for Dokumentation*, vol. 44, no. 1, 27-31

Bales, Kathleen and Alan Tucker. (1988) The RLIN Database: Current Status, Work in Progress, Future Development. *Cataloging and Classification Quarterly,* v. 8, no. 3, 79-89

Barry, Walfish. (1983) Hebrew and Yiddish Personal Name Authorities under AACR2. *Cataloging and Classification Quarterly,* vol. 3, no. 4, 51-64

Blanken, Robert R. (1971) The Preparation of International Author Indexes, With Particular Reference to the Problems of Transliteration, Prefixes, and Compound Family Names. *Journal of the American Society for Information Science*, vol. 22, no. 1, 51-63

Bossmeyer, Christine and Stephen W. Massil, ed. (1987). *Automated Systems for Access to Multilingual and Multiscript Library Materials—Problems and Solutions: Papers from the Pre-conference Held at Nihon Daihaku Kaikan Tokyo, Japan, August 21-22, 1986.* Edited for the Section on Library Services to Multicultural Populations and the Section on Information Technology. West Germany: Saur.

Breeding, Marshall. (2002) Association of Research Libraries: Current Automation Systems. Retrieved 10/13/2002 from http://www.librarytechnology.org/arl.pl

Butcher, Roger. (1993) Multi-lingual OPAC Development in the British Library. *Program,* vol. 27, no. 2, 165-171

Byrum, John D., Jr. and Olivia Madison, ed. (1998). *Multi-script, Multilingual, Multi-character Issues for the Online Environment: Proceedings of a Workshop Sponsored by the IFLA Section on Cataloguing, Istanbul, Turkey 24, 1995*. West Germany: Saur.

CARL Shows Multilingual Interfaces. (1998) *Advanced Technology Libraries,* vol. 27, no. 8, 5

Chakraborty Ajoy Ranjan; Chakrabortty Bhubaneswar. (1976) On Transcription and Transliteration. *Annals of Library Science and Documentation,* vol. 23, no. 3, 238-241

Chepesiuk, Ron. (1997) Expanding Internationally: OCLC Gears Up. *American Libraries,* vol. 28, no. 10, 52-55

Clavel-Merrin, Genevieve. (1999) The Need for Co-operation in Creating and Maintaining Multilingual Subject Authority Files. *65th IFLA Council and General Conference. Bangkok, Thailand Aug. 20-28, 1999*. Retrieved 12/9/2000 from http://www.ifla.org.sg/IV/ifla65/papers/080-155e.htm

Delsey, Tom. (2000) The Library Catalogue in a Networked Environment. Paper Written for the *Bicentennial Conference on Bibliographic Control for the New Millennium*. Library of Congress, Washington, D.C., November 15-17, 2000. Retrieved 5/2/2002 from http://lcweb.loc.gov/catdir/bibcontrol/delsey_paper.html

Dorman, David. (2002) OCLC Taps VTLS to Help Migrate WorldCat. *American Libraries*, vol. 33, no. 5, 83

Durance, Cynthia J. (1987) What Next? : Issues Arising from Conference Deliberations. (185-191) In Bossmeyer, Christine and Stephen W. Massil, ed. (1987). Automated Systems for Access to Multilingual and Multiscript Library Materials—Problems and Solutions ...

Eilts, John. (1996) Non-Roman Script Materials in North American Libraries: Automation and International Exchange. *International Cataloguing and Bibliographic Control,* vol. 25, no. 3, 51-53

Elman, Sarah Su-erh. (1991) Automation in U. S. East Asian Libraries in the United States: A Review and Assessment. *College & Research Libraries,* vol. 52, no. 6, 559-73

Elrod, McRee. (1980) Non-Roman Records and Automated Cataloguing. *International Association of Orientalist Librarians Bulletin,* vol. 17, 13-17

Erickson, Janet. (1997) Options for Presentation of Multilingual Text: Use of the Unicode Standard. *Library Hi Tech,* vol. 15, no. 3-4, 172

Fayen, Emily Gallup. (1989) The ALA Character Set and Other Solutions for Processing the World's Information. *Library Technology Reports,* vol. 25, no. 2, 255-273

Fung, M. C. (1983) *Development in the Computerization of Non-Roman Scripts*. Singapore: Maruzen Asia.

Gorman, Michael and Paul W. Winkler, ed. (1998) *Anglo-American Cataloging Rules: Prepared under the direction of the Joint Steering Committee for Revision of AACR2*, *2nd. Ed. 1998 revision*. Chicago: American Library Association.

Guidelines for Multilingual Materials Collection and Development and Library Services. (1990) Prepared by the Multilingual Materials Subcommittee (ad hoc), Adult Library Materials Committee, Reference and Adult Services Division, American Library Association. Adopted by the Reference and Adult Services Division Board of Directors. Retrieved 5/14/2002 from http://www.ala.org/rusa/stnd_multiling.html

Holt, Brian P.; Sally H. McCallum and A. B. Long (1987) *UNIMARC Manual*. London: International Federation of Library Associations and Institutions, Universal Bibliographic Control and International MARC Programme, British Library Bibliographic Services).

INNOPAC Millennium Non-roman Character Sets Support Arabic. (1997) *Information Today,* vol.14, no.10, 51

Iskandarani, A. I. and M. A. Anwar. (1992) Automated Bilingual Circulation System Using PC Local Area Networks. *Microcomputers for Information Management,* vol. 9, no. 3, 161-176

Jover, B. (1987) *The Review of the International Standard Bibliographic Descriptions (ISBDs) and Non-Roman Script Publications* (IFLA Publications, no. 38). New York, NY: K.G. Saur. 205-212

Karen, Wei. (1988) Computer Applications in East Asian Libraries: Current Status and Future Trends," *International Symposium on New Techniques and Applications in Libraries*, Xi'an, Chinese People's Republic, Sep. 8-11. Sponsored by Xi'an Jiao Tong University (Chinese People's Republic), Ohio University (USA), and Kanazawa Institute of Technology (Japan), and published by Xi'an Jiao Tong University Press. 456-462

Khurshid, Zahiruddin. (1992) Arabic Online Catalog. *Information Technology and Libraries*, vol. 11, no. 3, 244-51

Khurshid, Zahiruddin. (1997) DMARC: An Extended USMARC Format. *Technical Services Quarterly,* vol. 14, no. 4, 37-46

Khurshid, Zahiruddin. (1998) ARABMARC: A Long Way to Go. *World Libraries,* vol. 8, no. 2, 60-68

Kim, Heesop et al. (1999) Correlation between users' characteristics and preferred features of Web-based OPAC evaluation. *ETRI Journal,* vol. 21, no. 4, 83-93

Landry, Patricia. Multilingual Subject Access Project: Comparative Analysis of Titles Indexed Using LCSH, RAMEAU and SWD/RSWK. Retrieved 4/22/2002 from http://members.theglobe.com/peterdale/mlsumm.htm

Lazinger, Susan and Judith Levi. (1996) Multiple Non-Roman Scripts in Aleph—Israel's Research Library Network. *Library Hi Tech,* vol. 14, no. 1, 111-116

LC to Freeze Card Catalog. (1977) *Library of Congress Information Bulletin,* vol. 36, no. 44, 743-746

Lee C. F. (1981) The ISBDs and Chinese Cataloguing, and their Application in Hong Kong Libraries. *International Cataloguing,* vol. 10, no. 1, 10-12

Lee, Thomas H. (1988) The Development of CJK Bibliographic Databases in North America and East Asia. *Cataloging and Classification Quarterly,* vol. 8, no. 3-4, 111-26

Leggott, Mark. (1991) Unique, Universal, and Uniform Character Encoding," *Canadian Library Journal,* vol. 48, no. 5, 345-46

Mariam, Abdul Kadir. (1981) ISBD: a Review of its Application in the Malaysian National Bibliography. *International Cataloguing,* vol. 10, no. 2, 22-24

Maruyama, Shojiro. (1987) Japanese Bibliographic Information—its Control and Standardization. *The Electronic Library,* vol. 5, no. 1, 44-48

McBurney, Donald H. (2001) *Research Methods*. Belmont, CA: Wadsworth/Thomson Learning. 246

McCallum, Sally and Monica Ertel, ed. (1995). *Automated Systems for Access to Multilingual and Multiscript Library Materials: Proceedings of the Second IFLA Satellite Meeting, Madrid, August 18-19, 1993*. Edited for the Section on Information Technology, the Section on Library Services to Multicultural Populations, and the Section on Cataloguing. West Germany: Saur.

Medawar, Katia. (1999) The Implementation of the Arabic Script in OLIB7 at the American University of Beirut Libraries. *Program*, vol. 33, no. 4, 303-312

Murtomaa, Eeva and Eugenie Grieg. (1994) Problems and Prospects of Linking Various Single-language and/or Multi-Language Name Authority Files. *International Cataloguing and Bibliographic Control,* vol. 23, no. 3, 55-58

Pasterczyk, Catherine E. (1985) Russian Transliteration Variations for Researchers. *Database,* vol. 8, no. 1, 68-75

Petzold, Charles. (1993) Move Over, ASCII! Unicode is Here," *PC Magazine,* vol. 12, (Oct.), 374; Viewing a Unicode TrueType Font under Windows NT. *PC Magazine,* vol. 12, (Nov.), 375; Typing Unicode Characters from the Keyboard," *PC Magazine,* vol. 12, (Dec.), 426

Powell, R. R. (1997). *Basic Research Methods for Librarians*. Greenwich, CT: Ablex.

Raptis, P. and Salaba A. (1994) Bilingual Authority Files at the Central Library of the Aristotle University of Thessaloniki, Greece. *International Information and Library Review,* vol. 26, no. 2, 67-76

Rogers, Gloria H. (1986) From Cards to Online: The Asian Connection. Information Technology and Libraries, vol. 5, no. 4, 280-284

Salant, Priscilla and Don A. Dillman. (1994) *How to Conduct your Own Survey*. New York: John Wiley.

Shaker, Ali (2002) Non-Roman script in Library OPACs: A Survey of Automated Library Systems. (Unpublished report)

Slater, M., ed. (1990). *Research Methods in Library and Information Studies*. London: The Library Association.

Simsova, Sylva (1985) Multilingual Information and the Computer. *Information Development*, vol. 1, no. 2, 103-08

Sirsi Delivers Unicorn 2002 (2002) *Sirsi Extra Newsletter,* vol. 1, no. 3, 5

Smith B; Smith V G I. (1976) Transliteration Systems for the Russian Cyrillic Alphabet: Report on a Survey. *International Cataloguing*, vol. 5, no. 2

Smith-Yoshimura, Karen. (2000) Original Scripts Debut in Eureka®. *RLG Focus,* vol. 47. Retrieved 5/1/2002 from http://www.rlg.org/r-focus/i47eureka.html

Sommer, F. E. (1934) Books in Foreign Script in the Public Library. *Library Journal*, vol. 59, no. 20, 892-93

Spalding, C Sumner. (1977) Romanization Reexamined. *Library Resources and Technical Services*, vol. 21, no. 1, 3-12

Steinberger, Naomi M. (1994) ALEPH: A Bilingual Integrated Library System. *Library Hi Tech,* vol. 12, no. 2, 93-96

Terekhova, L. A. (2000) Choice of Language of the Bibliographic Description in Multilingual Catalogues. *International Cataloguing and Bibliographic Control,* vol. 29, no. 3, 53-4

Tillett, Barbara. (2000a) Authority Control at the International Level. *Library Resources and Technical Services,* vol.44, no.3, 168-172

Tillett, Barbara B. (2000b) Authority Control on the Web. Paper Written for the *Bicentennial Conference on Bibliographic Control for the New Millennium*, Library of Congress, Washington, D.C., November 15-17, 2000. Retrieved 5/3/2002 from http://lcweb.loc.gov/catdir/bibcontrol/tillett_paper.html

Tyckoson, David. (1991) The Twenty-First Century Limited: Designing Catalogs for the Next Century. (p.25) In Tyckoson, David, ed. *Enhancing Access to Information: Designing Catalogs for the 21st Century*. (New York: The Haworth Press, Inc.)

Understanding MACR Bibliographic: Machine-Readable Cataloging. (2000) Written by Betty Furrie in conjunction with the Data Base Development Department of The Follett Software Company. Washington, DC: LC-Cataloging Distribution Service.

The Unicode Consortium. (1991-92) *The Unicode Standard: Worldwide Character Encoding, Version 1.0* (Reading, Mass.: Addison-Wesley); supplemented by *The Unicode Standard, Version 1.1* (prepublication ed.) (Mountain View, Calif.: Unicode Consortium, 1993. (Unicode is a trademark of Unicode, Inc.)

USMARC Specifications for Record Structure, Character Sets, Tapes. (1990) Prepared by Network Development and MARC Standards Office (Washington, D.C: Cataloging Distribution Service, Library of Congress).

Vassie, Roderic. (1998) Marrying the Arabic and Latin Scripts Conceptually. *International Forum on Information and Documentation*, vol. 23, no. 3, 20-27

Vernon, Elizabeth. (1991) Hebrew and Arabic Script Materials in the Automated Library: The United States Scene. *Cataloguing and Classification Quarterly,* vol. 14, no. 1, 49-67

Vernon, Elizabeth. (1996) Decision-Making for Automation: Hebrew and Arabic Script Materials in the Automated Library. *Occasional Papers*, Number 205. University of Illinois, Urbana. Graduate School of Library and Information Science: 96p. (ERIC Document Reproduction Service No. ED398-911)

Wei, Karen T. and Noguchi, Sachie (1989) RLIN CJK versus OCLC CJK: the Illinois Experience. *Library Resources & Technical Services*, vol. 33, no. 2. 140-151. In Zeng, Lei. (1991) Automation of Bibliographic Control...

Weinberg, Bella (1974) Transliteration in Documentation. *Journal of Documentation,* vol. 30, no. 1, 18-31

Wellisch, Hans Hanan. (1976) Script Conversion Practices in the World's Libraries. *International Library Review,* vol. 8, no. 1, 55-84

Wellisch, Hans Hanan. (1978) Script Conversion and Bibliographic Control of Documents in Dissimilar Scripts: Problems and Alternatives. *International Library Review, v*ol. 10, no. 1, 3-22

Wellisch, Hans Hanan. (1980) The Exchange of Bibliographic Data in Non-Roman Scripts. UNESCO *Journal of Information Science, Librarianship and Archives Administration, v*ol. 2, no. 1, 13-21

Why Unicode? (2000) *Library Systems,* vol. 20, no. 10, 79-80

Wong, William Sheh. (1992) Automation and Library Administration with Special Reference to US East Asian Collection. *Journal of Educational Media and Library Sciences, v*ol. 29, no. 2, 113-26

Young, Heartsill, ed. (1983) *The ALA Glossary of Library and Information Science*. Chicago: American Library Association. 245

Young, Joann S. (1992) Chinese Romanization Change: A Study on User Preference. *Cataloging & Classification Quarterly,* vol. 15, no. 2, 15-35

Zeng, Lei. (1991) Automation of Bibliographic Control for Chinese Materials in the United States. *International Library Review, v*ol. 23, 299-319

Zeng, Lei. (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)