



BOBCATSSS 2010 @ Parma, Italy

Dates: Monday 25th, Tuesday 26th, Wednesday 27th January, 2010

**Bridging the digital divide:
libraries providing access for all?**

Bibliographic Control of Foreign Languages: the Case of Displaying Cyrillic Characters in Online Catalog at the University of Florida

Alena Aissing, Associate University Librarian, University of Florida

Introduction

Libraries hold foreign language collections that include non-Roman alphabets (e.g., Cyrillic, Arabic, Hebrew, Greek, Japanese, etc.). Students are faced with challenges when they wish to access these publications due to transliteration schemes that have been used for many decades. While there are various different transliteration systems in the world, the American libraries use the Library of Congress system. The libraries' integrated systems could only display the Romanized text in bibliographic records. This creates barriers for students to access, for example, Russian publications in online catalogs. Today, there is a new universal multi-script character set, International Standard ISO/IEC 10646 – the Unicode. The Unicode Consortium, the nonprofit organization that coordinates Unicode's development, has the ambitious goal of eventually replacing existing character encoding schemes with Unicode and its standard Unicode Transformation Format (UTF) schemes, as many of the existing schemes are limited in size and scope and are incompatible with multilingual environments.¹ This standard, if implemented in the library system; enables not only to convert Romanized bibliographic data into a linguistically accurate display in the original alphabet but also users would be able to search and print in non-Roman scripts. Libraries are slowly implementing systems based on this new standard. There are some crucial issues that involve Unicode implementation in the libraries. What are the solutions for multiple languages in library systems? The University of Florida holds collections in several non-Roman scripts. As Slavic Studies specialist, my attention has focused on integrating records for Russian into the main OPAC, which now includes over 10,000 bibliographic records for monographs in original Russian alphabet.

Academic libraries continue to purchase or subscribe to materials in foreign languages which necessitate foreign language skills. A 2004 study done by Ms. Kellsey and Kievel from the University of Colorado Libraries revealed that humanities scholars continue to use sources in the same level from forty years ago.² The authors counted 16, 138 citations within 468 articles found in four journals from history, classics, linguistics, and philosophy in the years 1962, 1972, 1982, 1992, and 2002 in order to identify trends in foreign-language citation behavior of humanities scholars over time. The number of foreign-language sources cited in the four subjects has not declined over time. Consistent levels of foreign-language citation from humanities scholars indicate a need for U.S. research libraries to continue to purchase these materials and to recruit catalogers and collection development specialists with foreign-language knowledge.

1 Unicode Consortium: <http://unicode.org>

2 Kellsey, Charlene, Knieval, Jeniffer E. (2004) Global English in the Humanities? A Longitudinal Citation Study of Foreign Language Use by Humanities Scholars. *College & Research Libraries*. 65 (3) : 194-204

The knowledge of foreign languages plays a significant role in most major functions of academic librarians. A 2008³ survey of academic librarians found that foreign languages are used mainly for selecting and evaluating material (47.8%), cataloging materials (38.2%) and communicating with patrons (32.8%).

Transliteration of non-Roman Scripts

Using university libraries can be overwhelming for many international students. The communication and cultural differences prevent many from effectively using the library. Librarians need to understand the communication barriers these students face. A 1999 study of non-native English-speaking students found that library technology was a leading cause of anxiety among library users.⁴ The act of finding, retrieving, and using library materials via online catalogs and databases as well as printing or photocopying articles may be daunting task for students without adequate computer education; and librarians need to be prepared to guide international students through these activities.⁵

Also American students in foreign language classes usually experience various difficulties in finding library materials in the language they study. In Russian, bibliographic records have been modified (i.e., Romanized) by transliteration into the Roman (English) alphabet. There are several different transliteration systems for Cyrillic script used throughout the world. Because the number of Roman letters in the English alphabet is less than the number of Cyrillic letters in the Russian alphabet, most of the transliteration systems have to resort to the use of diacritics or letter combinations to achieve complete transliteration. This can present several problems for the users who have to deduce how these letter combinations translate back to the original Cyrillic alphabet. For example Russian letter **ц** is represented with two letters **ts** in the Roman alphabet. In addition, Russian alphabet has no **h** and represent this sound mostly as **Г** (g in English), therefore transliterating Hamlet from Cyrillic back to roman script results in Gamlet. A Russian name beginning with **Я** could be transliterated into ia, ja, or ya. This causes inconsistencies in international transliteration schemes that lead to major retrieval problems.

In the early 1990s I conducted a study at three major universities; University of Florida, Florida State University and University of Illinois with 145 undergraduate and graduate

3 Zhang, Li. (2008) Foreign Language Skills and Academic Library Job Announcements: A Survey and Trends Analysis, 1966-2006. *The Journal of Academic Librarianship*. 34 (4) : 322-31

4 Agada, John. Inner-City Gate Keepers: An exploratory Survey of their Information Use Environment. *Journal of the American Society for Information Science*. 50 (1) : 74-85

5 Baron, S. and Strout-Dapaz, A. (2001) Communicating with and Empowering International Students with a Library Skills Set. *Reference Service Review*. 29 (4) : 314-326

students.⁶ The randomly selected students were tested by the use of three specially designed tests and a questionnaire. **Test A** intended to investigate how successful the students would be in searching and retrieving transliterated records without any instructions. **Test B** was given to students after they got instruction on how to use the Library of Congress transliteration table for Russian. After instruction, **test C** included a list of transliterated Russian titles and author's names. If a student were to search a book written in Cyrillic script in the online catalog, this would be the way in which it would display. The user then would have to match the transliterated information with its Cyrillic version and reestablish the text in the original version to determine if a given record matches the one sought for.

The results were very interesting. In **test A**, where no instructions were given, students were asked to provide searches for bibliographic records based on their own concept of transliteration. They had to reach to their knowledge of Russian phonology and orthography. Out of 145 students who took test A, 14 partially translated the Russian words into English instead of transliterating them. There were 93 letters (19 words) in test A. Commonly encountered authors such as **Александр Сергеевич Пушкин** (Aleksandr Sergeevich Pushkin) were used. The largest possible number of mistakes was 93, since there were 93 letters. Test results showed without knowing the transliteration table, none of the students would be able to conduct a 100 percent successful search. The lowest number of mistakes was one and only one person achieved that rate. The average number of mistakes made was 21.6. The largest group of students, 28 made between 11 and 15 mistakes.

Test B – testing students after instructions – showed that overall improvement. Twenty students transliterated all 82 characters without any mistakes followed by 59 students who made fewer than 6 mistakes. This constitutes the largest student unit, forming 43 percent of the total sample. Only one person had all 82 letters wrong. The average number of mistakes made 8.9.

Test C showed that after instruction, three students did not make any mistakes in reversing the transliteration process. The largest group consisting of 20 students, made 3 mistakes while only 5 students made as many as 16 to 17 mistakes. Another factor examined in test C was the total number of mistakes that students made individually for each Russian letter. The most misunderstood letter was **И** (140), followed by the letters **Я** (96), **Ы** (94), **Ю** (90), and **Й** (80).

The intent of this study was to examine the public reaction to online retrieval of material involving Cyrillic alphabet transliterated into Roman letter in a primarily English environment. As expected, the statistical analysis strongly suggested that the use of transliteration in retrieving bibliographic records forms a barrier to access non-Roman scripts even for those skilled in the original script. The important point is that even after hours of instructions in the LC tables; 105 students (72 percent) still felt (according to the given questionnaire) that searching in the original alphabet would be more efficient and easier. It is likely that such conclusion would apply to records in other non-Roman scripts.

Unicode

“The emergence of the Unicode Standard, and the availability of tools supporting it, is among the most significant recent global software technology trends.”

Unicode Consortium (<http://www.unicode.org/standard/WhatIsUnicode.html>)

Libraries have been providing access to non-roman scripts for the past 25 years, yet the libraries' integrated library system only displayed the Romanized bibliographic records. Since 1991, a universal character set for encoding and displaying the non-English vernacular scripts through Unicode standard in the libraries, is available. Unicode is a standard for a universal characters set for encoding the characters in the scripts of the world's languages. It is fully compatible 16-bit version of an international standard developed by the International Organization for Standardization (ISO) and the International Electrotechnical Committee (IEC), known as ISO/IEC 10646 or the Universal Multiple-Octet Coded Character Set. The Unicode standard was developed by a consortium of interested parties, known as the Unicode Consortium. The members of the organization are computer businesses such as Apple Computer, IBM, Microsoft, Adobe and some library systems companies like Online Computer Library Center (OCLC) and Research Libraries Group (RLG). Unicode was approved for encoding MARC records in 1998. A number of factors have come together in recent years to contribute to the usability of Unicode for the libraries. MARC21 specifications now include support for the use of the Universal Coded Character Set (ISO/IEC 10646) for exchange of MARC21 records.⁷

Unicode and Libraries

Many factors have to come together for libraries and system vendors in order to implement Unicode standard. Although it has been more than a decade since the Unicode standard was developed, academic libraries are slow in implementing it to their online catalogs. Implementation of this sixteen-bit character encoding that can represent the principal written languages collected by American libraries means a revolutionary change in serving foreign students. This standard is important as for the libraries will be able to deliver linguistically accurate display to users who seek to access non-roman scripts.⁸

Project at the University of Florida Libraries

The library integrated system Ex Libris (USA) Inc. joined the Unicode Consortium in 1999. The Consortium brings together leading industry corporations such as Microsoft, ORACLE, and SUN and includes researchers at the leading edge of emerging international character encoding standards.

7 Tull, Laura. (2002) Library Systems and Unicode: A Review of the Current State of Development. *Information Technology and Libraries*. 21 (4) : 181-185.

8 Coyle, Karen. (2005) Unicode: The Universal Character Set. *The Journal of Academic Librarianship*. 31 (6) : 590-592

At the University of Florida, the Aleph® integrated library system currently uses UNICODE (UTF8) internally which is one reason why it is a multi-lingual product capable to offers support of over 20 languages and scripts including bi-directional display capabilities and we're adding CJK.

With assistance of my colleague at the University of Florida, cataloger Librarian for Chinese, Japanese and Korean, Ms. Hikaru Nakano, a project was launched in 2009 to add Cyrillic characters to 10,000 existing MARC records, using Latin2Cyrillic, a transliteration macro developed by Joel Hahn, Niles Public Library District in Niles, Illinois which was used with the OCLC interface.⁹

We also researched and followed an example of the Queens Borough Public Library, which was one of the first libraries to have implemented Cyrillic in 2003 in their online catalog by re-transliterating their Romanized records back into the original Cyrillic alphabet.¹⁰

In an informal inquiry put on the Slavic Librarians' Listserve by Ms. Tatiana Barr, Cataloger Librarian/Team Leader at Yale University Library indicated that whereas many libraries began the conversion in 2008, others such as University of Pittsburg and University of Chicago have been working on similar project for a number of years.¹¹

We successfully requested funding via the University George A. Smathers Libraries Mini Grant for two OPS students to complete quality control that measured the correctness of Russian bibliographic records after implementing the project.

The successful completions of this project lead to inclusion of Cyrillic characters to approximately 10,000 bibliographic records. These records are shared with the Florida State University System as well as with all OCLC member libraries.

Conclusion

With Unicode implementation into library management system the integral support to display multi-scripts for non-Roman languages is possible. There are many issues that still need attention: authority control, subject access, input methods and fonts, keyboards, web browser display, inclusion of journal titles, etc.

There is a need to conduct surveys with students of Russian to assess user satisfaction and improvement to access to original Russian bibliographic records in the OPAC. There is also necessity for training for both librarians and patrons. Before training, outreach activities and publicity will be necessary.

⁹ ACRL. Slavic Cataloging Manual. OCLC Macros for Cyrillic Parallel Fields:
<http://www.indiana.edu/~libslav/slavcatman/cyrillicmacros.html>

¹⁰ Jacobs, Jane W. and Das, Malabika. (2005) Making the Cyrillic OPAC a Reality. Slavic and East European Information Resources. 6 (2) : 135-149

¹¹ Personal e-mail exchange with Ms. Tatiana Barr, Cataloger Librarian, Yale University.

Ms. Nakano and I will publish the description of our completed project in the near future.

Alena Aissing: European Studies Associate Librarian, University of Florida George A. Smathers
Libraries, Gainesville, Florida. USA