

Тенденции применения транслитерации и UNICODE в российских библиотеках

The Trends in Using Transliteration and UNICODE in Russian Libraries

Тенденції застосування транслітерації й UNICODE в російських бібліотеках

М. Л. Халабия

Российская государственная библиотека, Москва, Россия

Mariya Khalabiya

Russian State Library, Moscow, Russia

М. Л. Халабія

Російська державна бібліотека, Москва, Росія

В докладе раскрываются основные тенденции применения транслитерации и UNICODE в российских библиотеках. Описываются проблемы, связанные с применением конверсии систем письма для представления библиографической информации в многоязычной электронной среде.

The main trends in using transliteration and UNICODE in Russian libraries are considered. The problems in conversion of lettering systems for presenting bibliographic information in the multi-language electronic environment are described.

В доповіді розглядаються основні тенденції застосування транслітерації та UNICODE в російських бібліотеках. Описуються проблеми, пов'язані із застосуванням конверсії систем письма для представлення бібліографічної інформації в багатомовному електронному середовищі.

В условиях интенсивной международной коммуникации и международного библиотечно-информационного сотрудничества появляется необходимость представить текст, написанный в одной графической системе средствами письменности другого языка. Этот процесс называется транслитерацией. В соответствии с ГОСТ 7. 79-2000 (ИСО 9-95) «Правила транслитерации кирилловского письма латинским алфавитом», под транслитерацией понимают конверсию систем письма, при которой каждый графический элемент представляется (заменяется) одним и тем же графическим элементом другой системы письма [1].

Поскольку в мире транслитерация в латиницу используется достаточно широко, а UNICODE не получил достаточного распространения, перед российскими библиотеками встала проблема адекватного представления кириллицы в процессе международного обмена [2]. Давайте рассмотрим тенденции конверсии кириллической системы письма в латиницу.

Применение транскрипции

Большинство российских библиотек при конверсии систем письма применяют практическую транскрипцию, под которой понимают передачу слова с одного языка на другой, при которой максимально сохраняется фонетический облик слова [3]. Прежде всего, она применяется при создании нормативных (авторитетных) файлов имен лиц. Такой способ в какой-то мере позволяет адекватно представить точное название имени собственного.

Применение «английской» транслитерации

Данный способ транслитерации стремится получить запись с использованием латиницы, которая при чтении по правилам английского языка давала бы название, отражающее исходное русское имя собственное. При этом один знак русского алфавита заменяется последовательностью двух или более латинских букв [4]. Однако, одна из сложностей такой конверсии заключается в самой структуре русского алфавита. Так, при фонетическом и орфоэпическом анализе символов алфавита выясняется, что русский алфавит состоит не из 33, а 42 букв, тогда как в латинице всего 26 букв

[5]. Тем не менее, на принципах «английской» транслитерации построены системы транслитерации британской библиотеки и Библиотеки Конгресса США, а также ГОСТ 7. 79-2000 (ИСО 9-95).

Применение специальной программы для ввода букв с диакритическими знаками

Ряд российских библиотек применяет программное обеспечение при создании многоязычной библиографической записи. Так, во Всероссийской государственной библиотеке иностранной литературы имени М. Рудомино уже много лет используется программа TINLIB, которая позволяет создавать библиографические записи практически на любом европейском языке, с использованием текстового редактора ЛЕКСИКОН. По заданию отдела автоматизации ВГБИЛ фирма «Микроинформ» создала набор шрифтов, состоящий из недостающих в стандартных наборах латиницы и кириллицы европейских языковых символов и символов с диакритикой. Там же создали драйверы для печати этих символов. Дополнительный набор символов разместили на вспомогательном поле, откуда каждый символ указателем мышки можно считать, нажав правую кнопку. Выбранный символ перемещается на первую, рабочую страницу ЛЕКСИКОНА, на место указанное курсором. На первой странице оператор с помощью клавиатуры готовит библиографическую запись каталожной карточки на кириллице, либо латинице, считывая по необходимости недостающие символы мышкой. Вот уже более десяти лет этот процесс применяется не только для создания библиографических записей, но и любых многоязычных текстов [6].

Применение UNICODE

UNICODE должен снять проблему представления символов различных алфавитов в одной системе и проблему международного обмена библиографическими данными. Поэтому ряд крупных библиотек России рассматривают возможность полноценного использования универсальной системы кодирования знаков.

В то же время, продукция компьютерной индустрии, включая автоматизированные библиотечно-информационные системы (АБИС), продается по всему миру. Производители таких систем вынуждены признавать необходимость поддержки национальных алфавитов и иных знаков.

Российская государственная библиотека, учитывая такую тенденцию, использует UNICODE в АИБС РГБ, основанной на программном обеспечении АЛЕФ, и проводит исследование соотношения символов алфавитов национальных языков России с UNICODE.

Под UNICODE понимают 16-разрядную систему кодирования, совместимую с системой ASCII, которая охватывает символы разных письменностей: латинской, кириллической, греческой и т. д., а также языков, использующих иероглифы [7]. UNICODE как универсальная и эффективная компьютерная система кодирования имеет дело с каждым символом как с уникальным значением и [8] базируется на десяти основных принципах:

- универсальность;
- эффективность;
- кодирование знака, но не образа символа в его орфоэпическом звучании
- семантика;
- ясный, понятный текст
- логический порядок;
- унификация;
- динамическая композиция;
- эквивалентное согласование;
- конвертируемость.

Таким образом, исходя из вышеперечисленных принципов функционирования UNICODE, критериями анализа для соотношения знаков алфавитов национальных языков России с UNICODE являются структура алфавитов и система кодирования UNICODE.

Основной целью при проведении данного исследования является усовершенствование поиска информации в электронном каталоге библиотеки, решение проблемы многоязычия при создании библиографической записи и, как следствие, – внесение предложений в Консорциум UNICODE по внесению ряда символов в таблицы «расширенной кириллицы» и «пунктуации». Основной

задачей исследования было стремление определить, что мы анализируем: орфоэпический и фонетический состав алфавитов национальных языков России с UNICODE или же только символы без их фонетических особенностей.

Так, кириллические гласные «А», «Е», «Ё», «И», «О», «У», «Ы», «Э», «Ю», «Я» практически во всех национальных языках звучат одинаково, исключение составляет кабардинский язык, в котором «Е» – йэ, «И» – иы, «О» – Ioу, У – уы. Остальные гласные символы алфавита кабардинского языка соответствуют звучанию русского языка. Кириллический символ «Г» фонетически представлен как «гэ» в русском языке, «гы» – в кабардинском языке, «гэ» – в чеченском, «ге» – в ногайском языке.

Фонетическую окраску символов алфавитов национальных языков России будем называть глифом, под которым понимают представление символа в побитовом отображении информации. Другими словами, глиф представляет собой визуальное и фонетическое отображение знака независимо от формы написания и звучания буквы в контексте национального языка. Так как основной особенностью универсальной системы кодирования знаков UNICODE является кодировка символа, не зависящая от его визуального представления и его различных способов правописания, то мы анализировали только символы расширенной кириллицы с UNICODE.

Все вышеперечисленные тенденции касаются и других российских библиотек. В подтверждение своих слов хотелось бы процитировать Вильяма Армса: «Библиотеки использовали широкий набор алфавитов задолго до специальных исследований по скриптам. В результате существуют огромные каталоги в формате MARC (и других форматах), содержащие кодировки в других системах, помимо UNICODE. Сейчас стоит устрашающая задача конвертации или обеспечения совместимости» [9]. Однако будущее, безусловно, за применением универсальной системы кодирования знаков UNICODE. Библиотеки должны скоординировать свои усилия в этом направлении, так как это позволит решить проблему передачи данных в многоязычной коммуникации.

Литература

1. ГОСТ 7. 79-2000 (ИСО 9-95) Правила транслитерации кирилловского письма латинским алфавитом: Офиц. изд. / Межгос. совет по стандартизации, метрологии и сертификации. – Мн: Межгос. совет по стандартизации, метрологии и сертификации, 2002. – 19 с. [1].
2. Халабия М. Л. Некоторые проблемы транслитерации и кодировки кириллицы // Информационные технологии, компьютерные системы и издательская продукция для библиотек: Материалы 10-й междунар. конф. LIBCOM – 2006. Электронный ресурс: <http://www.gpntb.ru/libcom6/disk/56.pdf> [2].
3. Белоозеров В. Н., Радковский Г. Н., Косарская Ю. П. Практическая транслитерация русского языка латинским алфавитом / В. Н. Белоозеров, Г. Н. Радковский, Ю. П. Косарская // НТИ. Сер. 2 Информ. процессы и системы. – 1997. – № 12. – с. 28–34. [3].
4. Белоозеров В. Н., Кузнецова А. Б. Транслитерация для банковских расчетов / В. Н. Белоозеров, А. Б. Кузнецова // НТИ. Сер. 1 Организация и методика информ. работы. – 2004. – № 10. – с. 28–32 [4].
5. Горбатков К. К проблеме латинской транслитерации кириллицы [5].
6. Страница отдела автоматизации ВГБИЛ: [электронный ресурс] http://autodpt.libfl.ru/home_autodpt/Osnova.htm [6].
7. Воройский Ф. С. Информатика. Энциклопедический систематизированный словарь- справочник: введение в современные информ. и телекоммуникационные технологии в терминах и фактах / Ф. С. Воройский. – М.: Физматлит, 2006. – 945 с. [7].
8. Laura Tull. Unicode // Encyclopedia of Library and Information Science. – Marcel Dekker, 2004. – p. 383–391. [8].
9. Армс В. Электронные библиотеки: Учеб. пособие / [пер. с. англ. яз. Арнаутова С. А. – М.: ПИК ВИНТИ, [2001]. [9].