

# СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

УДК 003.349.3.034:003.344

А. Х. Султанов, Ш. Р. Низамов

## Опыт практической транслитерации русского текста латинским алфавитом

*Описывается проблема, связанная с множественностью кодировок русского языка в электронных сообщениях и способ их решения для иностранцев, желающих переписываться со своими русскими коллегами, но не имеющими возможность установить или настроить соответствующую программу русификации. Содержание предложения состоит в том, чтобы, используя буквы только латинского алфавита, максимально приблизить их к буквам русского алфавита. При этом за основу берется не близость по произношению, а сходство по написанию. Приведенная в статье кодировка использовалась автором при переписке с зарубежными студентами и показала свою эффективность при отсутствии других способов общения на русском языке.*

Ни для кого не секрет, что сеть Интернет является самой большой в мире сетью, объединяющей в единое информационное пространство миллионы компьютеров. Колоссальными темпами растет количество информационных ресурсов, доступных в сети [1]. Не следует забывать, что количество пользователей сети также удваивается в среднем каждые полгода. Для многих пользователей Интернет становится не только средством поиска необходимой информации, но и средством повысить уровень своего образования дистанционным способом. Также хорошо известно, что Интернет лучшее место для общения с другими пользователями со всего мира и, практически, на любом языке мира. Но многие сталкиваются с большими проблемами при поиске информации или общении на каком-либо другом языке, кроме того языка, который установлен на данном компьютере пользователя. Выходом из данной ситуации может служить установка локализованной для этого языка версии программного обеспечения, либо установка соответствующего программного пакета, обычно включающего шрифты национального языка. Однако существует целый ряд проблем, препятствующих развитию сети, одной из которых является множественность кодировок национальных алфавитов и, в частности, русского. Например, только кодировок арабского алфавита существует более 30, а кириллического алфавита — около 10 (широко используются только 5). Вопрос унификации кодировок неоднократно поднимался на различных конференциях, но пока еще далек от какого-либо решения. Русскоязычные серверы работают во всех широко используемых кодировках. То же самое относится и к сообщениям по электронной почте. Не секрет, что даже электронное письмо, посланное в одной кодировке, может быть испорчено почтовым сервером, не работающим с данной кодировкой. Случаются ситуации, когда письмо не может быть прочитано из-за особенностей настройки почтовой программы адресата.

Рассмотрим теперь ситуацию, когда один пользователь хочет передать текст, содержащий симво-

лы национального алфавита, своему иностранному коллеге. Возможно несколько способов. Один из способов решения этой проблемы заключается в передаче по сети не текста, а его графического образа, т. е., проще говоря, файла в одном из распространенных в сети формате. Достоинства данного способа в полной идентичности передаваемой информации покрываются его недостатками, такими, как большой объем передаваемой информации, использование низкоскоростных каналов передачи данных, невозможность редактирования полученного текста. К тому же не все программы поддерживают работу в графическом режиме. Таким образом, данный подход вряд ли можно считать приемлемым.

Другой способ решения проблемы представления символов заключается в передаче пользователю не самого текста, а его образа, подготовленного в специальных форматах таких, как PostScript или Adobe PDF. Однако и этот способ имеет те же недостатки, что и при передаче графического образа. Такие, как большой объем передаваемой информации, невозможность дальнейшего редактирования. Кроме того, на компьютере пользователя должно быть установлено соответствующее программное обеспечение, что не всегда возможно.

Существует еще одно решение данной проблемы, когда пользователь передает просто файл в формате plain text другому пользователю. В этом случае объем передаваемой информации эквивалентен самому тексту, однако для просмотра такого текста на компьютере пользователя уже должен быть установлен соответствующий шрифт с поддержкой символов национального алфавита.

Как же выходить из данной ситуации. Для российских пользователей данная проблема вполне разрешима самыми различными путями, поскольку можно заранее договориться, какая кодировка будет использоваться в электронной корреспонденции, или использовать многочисленные программы перекодировки писем.

Кроме того, давно существует стандарт [2] на

передачу русских сообщений латинским алфавитом, суть которого заключается в том, чтобы заменить кодировку букв локального алфавита на аналогичные английские, которые понимает любая программа чтения и отправки писем по электронной почте.

Тем не менее, существуют ситуации, когда использование подобного языка необходимо. Например, многие зарубежные студенты, изучающие русский язык, не имеют возможности для установки программного пакета поддержки русского языка на свой компьютер. Причин может быть несколько, начиная от недостаточной квалификации до отсутствия соответствующих прав пользователя на изменение программного обеспечения компьютера, или использование терминалов. В подобной ситуации, если иностранец желает общаться с русскоговорящим пользователем на русском языке, самый простой путь именно в замене русских букв на их английские эквиваленты.

В настоящей статье предлагается одна из подобных кодировок, которая использовалась при переписке с иностранными студентами, изучающими русский язык.

Необходимо заметить, что существующий стандарт не всегда является удобным в использовании, когда речь идет о простом общении или переписке. В этом случае гораздо эффективнее применять практическую транслитерацию. Различные авторы используют в качестве критерия оценки практичности различные принципы. Это может быть принцип исторической преемственности [3] или чтение по правилам какого-либо языка. В данном же случае ставилась задача создания алфавита по принципу графического подобия, что одновременно приводит и к удобочитаемости и к экономии ресурсов.

Исходя из данного принципа основной задачей ставилось создание алфавита, который как можно точнее имитирует буквы русского алфавита, что не во всех случаях оказалось возможным.

Результаты замены приведены в таблице:

Русский символ	Аналог	Русский символ	Аналог
а	A	р	R или P
б	цифра 5	с	C
в	B	т	T
г	G	у	Y
д	D	ф	F
е, ё	E	х	X
ж	J	ц	TC
з	цифра 3	ч	цифра 4
и	I	ш	W
й	I	щ	W
к	K	ь	'b (апостроф и B)
л	L	ы	b1 (B и цифра 1)
м	M	ь	B
н	N	э	цифра 3
о	O	ю	U
п	P	я	J или JA

Новым в данном подходе является использование заглавных букв английского алфавита, многие из которых более похожи на символы русского. Использование цифр и сочетаний букв и цифр позволило заменить некоторые специфические символы русского языка.

Конечно, данная таблица не лишена и недостатков, так как для передачи некоторых символов используются одни и те же буквы английского языка. Так, русская буква "р" и "п" может быть заменена латинской "P". В большинстве случаев из контекста понятно, какая буква используется в данном слове, и само слово может быть однозначно прочитано или может быть найдено по словарю. То же самое можно сказать относительно замены ряда букв латинской буквой "J" и "W". Тем не менее, данное обстоятельство приводит к тому, что в последующем невозможно будет создать компьютерную программу, однозначно кодирующую и декодирующую любое сообщение, записанное с помощью данного алфавита.

Другим очевидным неудобством подобного алфавита является набор такого текста. Как правило, руки привыкают к раскладке клавиатуры и с трудом переходят на другой способ набора текста. В результате на один и тот же текст затрачивается большее время при использовании приведенной выше кодировки. Справедливости ради стоит заметить, что данная проблема присуща не только этой, но и всем подобным кодировкам. Наиболее простым выходом из данного положения может быть написание небольшого макроса для распространенных текстовых редакторов, например MS Word, либо программы на любом языке программирования, перекодирующих текст из русского языка в псевдорусский.

Практический опыт использования данной кодировки при переписке на русском языке со студентами из таких разных по языку и образу мышления стран, как Япония и США, показал, что все с первого раза понимали суть замены и были в состоянии прочитать и правильно интерпретировать русский текст, записанный латинским алфавитом.

#### ПРИЛОЖЕНИЕ

(В качестве примера, по традиции [3], приведем пример транслитерации небольшой части аннотации данной статьи)

СТАТЬЯ РОСВJWЕНА ОPИСАНИУ ПРОБЛЕМЫ,  
СВJЗАННОИ С  
МНОЖЕСТВЕННОСТЬЮ КОДИРОВОК РУССКОГО  
JЗЫКА В  
ЭЛЕКТРОННЫХ СООБЩЕНИJХ, ОПИСЫВАЕТСJ  
СРОСО5 IX РЕШЕНИJ ДLJ  
ИНОСТРАНТСЕВ JELAUWIX PEREPICЬ1BATЬCJ  
СО СВОИМI РУССКИМI  
KOLLEGAMИ, НО НЕ ИМЕУWIX ВОЗМОЖНОСТИ  
УСТАНОВИТЬ ИLИ  
НАСТРОИТЬ СООТВЕТСТВYUWYU PROGRAMMY  
РУССИФИКАТСИ.

#### СПИСОК ЛИТЕРАТУРЫ

1. Чем измерить Internet // Computer Weekly.— 1998.— № 31.— С. 21.
2. ГОСТ 16876-71. Правила транслитерации букв кирилловского алфавита буквами латинского алфавита.— М.: Изд-во стандартов, 1981.
3. Белоозеров В. Н., Радковский Г. Н., Косарская Ю. П. Практическая транслитерация русского текста латинским алфавитом // НТИ. Сер. 2.— 1997.— № 12.— С. 28-34.

Материал поступил в редакцию 14.09.99.