

При реализации обращения к справочникам необходимо сделать работу максимально удобной, что было достигнуто оформлением их на страницах в виде таблиц с данными. При такой организации интерфейса работа с системой становится максимально простой и легкой в освоении даже для неспециализированного пользователя. Ввод материала в систему также максимально упрощен и позволяет осуществлять наполнение базы достаточно быстро.

Под инструментом проведения исследований понимается модуль, позволяющий формировать произвольные запросы с учетом всех свойств и значений данных. Этот инструмент имеет средства для упорядочивания, выборки, операций с выборками, для построения словников, для отображения результатов в различной форме. Основное его предназначение состоит в автоматизации картографирования лексики русского диалектного языка и динамического визуализирования результатов обработки материала на карте.

Разработанная технология позволяет пользователю при построении лингвогеографических карт самостоятельно определять как картографируемый материал, так и обозначения, которые будут использоваться на карте. Обработка материала проводится в несколько шагов, которые позволяют пользователю не только выбрать материал, но и удалить данные, которые не представляют интереса.

Результатом проделанной работы является создание лингвогеографической информационной системы с доступом к данным через Интернет. Данная разработка масштабируется на любые территории и на любые типы и объемы информации.

## Учет регионального компонента на примере элементов шорской культуры и программы MS WORD

О. А. Соседко

Кузбасская государственная педагогическая академия,  
Новокузнецк, Россия

Чаще всего разработчики программного обеспечения учитывают или популярные, или глобальные национальные приоритеты. В частности, русифицировано большинство программных продуктов для российских пользователей. Но мир языков многообразен. Например, в кузбасском регионе есть малая народность — шорцы, обладающая своей культурой и языком. Поэтому в настоящее время необходимо адаптировать программные продукты и для нужд представителей малых народов.

Чтобы понять, какие проблемы возникают перед пользователем-шорцем, нужно познакомиться с шорским алфавитом.

Современный шорский алфавит основан на кириллице, но имеет ряд специальных букв, в большинстве сходных с буквами тюркских алфавитов. Эти буквы имеют дополнительные элементы — «хвостики», «черточки» или «двоеточие» над буквой.

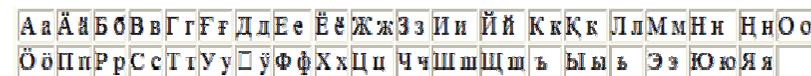


Рис. 1. Современный шорский алфавит

Как видно, проблемой является ввод спецбукв, так как для этих букв нет соответствующих клавиш на клавиатуре.

Проанализируем возможные решения проблемы.

1. Изменить наклейки тех клавиш, символы которых не используются в шорском языке. Перепрограммировать эти клавиши, чтобы при нажатии появлялись на экране эти спецсимволы. Иначе говоря, можно разработать по аналогии с русификатором программу «шорификатор». Недостаток этого способа состоит в том, что за

компьютером работает не один, а несколько пользователей и перелка клавиатуры будет неудобна «нешорским» пользователям. А «шорификатор» без нужных клавиш ненагляден и тоже неудобен.

2. Текст набирают в текстовых редакторах. Самым популярным из них, без сомнения, является MS Word. И самое простое, что доступно пользователю, это вставка символа.

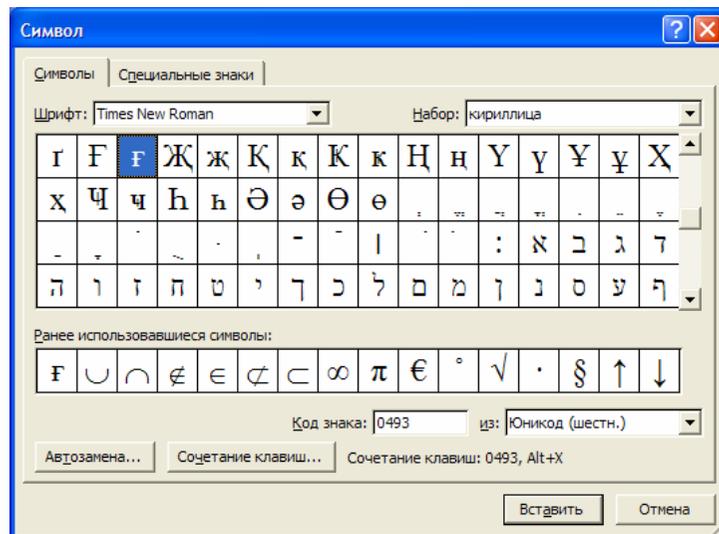


Рис. 2. Вид окна вставки символа

3. MS Word — это многофункциональный редактор, поэтому можно создать макросы, привязанные к кнопкам на панели инструментов. Это автоматизация второго способа. Лучше использовать программы Office 2000/2003, так как основные, самые популярные шрифты содержат расширенную кириллицу, в том числе особые шорские буквы. Достоинства этого способа состоят в удобстве ввода, хранения и печати. Недостатками являются: возможность хранения только в форматах DOC, RTF, HTML, необходимость сохранять настройки интерфейса в специальном DOT файле и переносить его на другой компьютер, необходимость учитывать совместимость разных версий, в противном случае произойдет потеря спецбукв или их замена на другие, в том числе латинские. На воспроизведение буквы на значке пиксельной матрицы после записи макроса нужно время.

4. В редакторе MS Word можно использовать автозамену, например, «Ғ» заменять на Ғ.

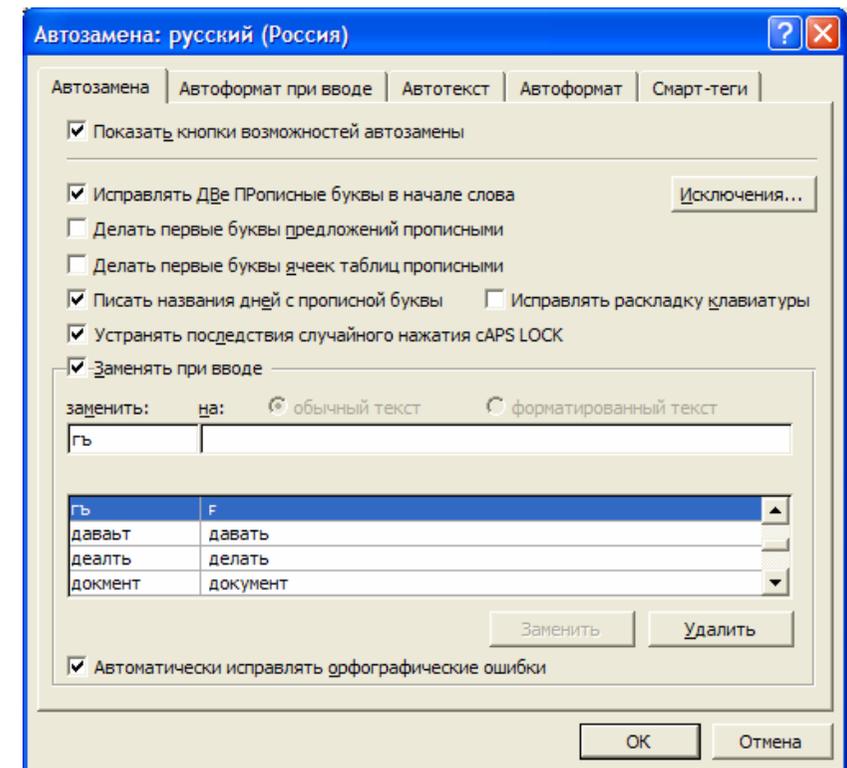


Рис. 3. Вид окна автозамены.

Явный недостаток этого способа в нерациональности: чтобы ввести один символ, нужно сделать четыре нажатия на клавиатуру: буква «Ғ», буква «ь», «пробел», «влево».

5. До появления программ Office 2000 специалисты по шорскому языку и другим тюркским языкам использовали другой подход. В конце двадцатого века был разработан специальный шрифт Siberia, в котором присутствовала часть шорского алфавита. Чтобы ввести символ, надо, как в первом способе, сделать вставку, выбрать шрифт Siberia, среди символов найти нужный, нажать «Вставить», закрыть окно вставки. Но при передаче файла

на компьютер, на котором не было Siberia, шорские буквы пропали.

6. В настоящее время разрабатывается и постоянно пополняется шрифт UNICODE, который в итоге должен содержать глифы символов любого языка мира. Шрифт UNICODE основан на четырехбайтовой кодировке, но это не значит, что каждая буква представляется четырьмя числами символов. Самые популярные, часто используемые в мире символы имеют меньше кодов для своей записи. Для латинского алфавита, как и прежде, используется однобайтовое кодирование, если его ASCII-код меньше 128. Затем идет большой раздел двухбайтового соответствия. В этом диапазоне находятся также полностью и русский, и шорский алфавиты.

Чтобы набрать букву этих алфавитов, надо знать соответствующие таблицы кодирования формата UTF8. Их можно узнать как в справочниках, так и в Интернете, например, по адресу <http://titus.uni-frankfurt.de/unicode/unitest.htm>. Каждый символ должен записываться двумя числами. Это можно делать с помощью цифровых клавиш на малой клавиатуре при нажатой клавише ALT. Например, букве «Ѣ» соответствует четырехзначный код 1171. Этот медленный способ и требует наличия на компьютере современного стандарта UNICODE UTF8, присутствующего в операционных системах Windows 2000, XP, 2003 и в их офисных программах.

Каждый из способов ввода символов имеет свои достоинства и недостатки, поэтому каждый пользователь решает, как вводить, редактировать и хранить национальный шорский текст<sup>1</sup>.

---

<sup>1</sup> Соседко 2003 — Соседко, О. А. Особенности обработки шорских текстов / О. А. Соседко // Повышение качества профессиональной подготовки будущего учителя информатики : сб. науч. тр. — Новокузнецк : КузГПА, 2003.

## Технологии создания информационной системы для работы с полнотекстовыми базами данных исторических документов<sup>1</sup>

В. О. Филатов, И. В. Кравцов  
Петрозаводский государственный университет, Россия

*Our article covers various terms and development technologies for web information systems for historical documents. The purpose of our system is to prepare electronic and printed editions for various kinds of medieval historical texts collections. We also describe a special research toolkit based on two original editors: SVG visual editor and XML markup editor. The article is built on logical pairs: needs of historians — our technological offers for their problems.*

Предлагаемая нами информационная система ориентирована на введение в научный оборот специализированного инструментария и методологий, позволяющих, во-первых, публиковать в сети Интернет комплексы средневековых документов, во-вторых, проводить на базе этих коллекций широкий спектр источниковедческих исследований и, в-третьих, организовать информационное пространство для совместной работы сообщества исследователей.

В данном тексте мы не ставим перед собой задачу аргументировать применение новейших компьютерных технологий в традиционных областях исторического источниковедения, так как это уже описано в других работах [Технология 2005: 241–281], кроме того, у каждого исследователя есть по этому поводу свое мнение. Мы лишь предлагаем платформу, с помощью которой можно это мнение подтвердить реальным экспериментом, перепроверить результаты этого эксперимента и подискутировать о методе.

На данный момент в систему вводятся документы двух комплексов источников: комплекса средневековых исторических источников «Moscovitica-Ruthenica» XII–XVIII вв. [Moscovitica

---

<sup>1</sup> Работа выполнена при финансовой поддержке РГНФ (проект № 06-01-12124в).