

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

ПРИНЦИПЫ И ЦЕЛИ ЭЛЕКТРОННОГО ИЗДАНИЯ ФОЛЬКЛОРНОГО АРХИВА БАШКИРСКОГО

ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

Б. В. Орехов, А. А. Галлямов, С.Ю. Данилин

Башкирский государственный университет, Уфа

The article deals with the digitalisation experience of the folklore archive of Bashkir State University. We can see either traditional features for folkloristics (full metainformation) either specific for machine presentation (replenishing index of words, interactive contents, geoinformation layer) in a public digital edition accessible through the Internet.

Очевидно, что в современных условиях хранение накопленного в фольклорных экспедициях материала на бумажных носителях — неэффективное решение. Осознав это, Башкирский государственный университет приступил к оцифровке своего фольклорного архива и его представлению в виде электронного научного издания, доступного в Интернете: <http://lcph.bashedu.ru/index.php?go=editions>. В первую очередь размещаются материалы на башкирском языке.

Электронное издание должно сохранять все сильные стороны бумажной публикации и использовать дополнительные возможности, которые даёт компьютерная форма. Таким образом, традиционный фольклористический минимум (аутентичность, полнота метайнформации) дополнен техническими решениями: автоматическая группировка и классификация материала, полный указатель словоформ с данными о частотности и гипертекстовой связью с конкордансом (ступень на пути к разного рода словарям). Представление текста тесно связано с геоинформационной составляющей проекта, не только позволяющей увидеть на карте место записи каждого текста, но и в будущем автоматически генерировать атласы распространённости мотивов, территориальной дистрибуции жанров и традиций.

Электронное научное издание «Фольклорный архив Башкирского государственного университета» — это постоянно пополняемая полнотекстовая сетевая многофункциональная коллекция фольклорных записей, собранных в Республике Башкортостан студентами и сотрудниками университета на протяжении последних пятидесяти лет.

Цели проекта:

1. Аккумуляция оцифрованных версий фольклорных произведений, собранных студентами и преподавателями БашГУ в фольклорных экспедициях с 1950-х годов по настоящее время;
2. Предоставление всем заинтересованным лицам свободного доступа к фольклорным архивам БашГУ;
3. Повышение эффективности научных исследований фольклорных произведений с помощью многофункциональной программно-информационной среды;
4. Популяризация произведений народного творчества.

ЭНИ предоставляет следующие пользовательские функциональные возможности:

1. Доступ пользователей к информации;
2. Поиск отдельных произведений по базе ЭНИ;
3. Лексический поиск во всех текстах, представленных в ЭНИ, с помощью указателя словоформ;
4. Работа со словарем частотности, связанным со страницами употребления словоформ;
5. Экспорт информации;
6. Оформление полноценных библиографических ссылок согласно ГОСТу;

Тексты в ЭНИ публикуются в том виде, в котором они зафиксированы в архиве, то есть никакой редакторской (и тем более литературной) правки, кроме исправления явных ошибок, не проводится. Во всех случаях сохраняются орфография и пунктуация источников. Важнейший критерий — точность и аутентичность передачи исходной текстовой информации. В существующем виде тексты публикуются (как в Интернете, так и вообще в доступном виде) впервые.

В ЭНИ используется двуязычная система. Тексты фольклорных произведений представлены только на языке оригинала, описания и элементы интерфейса — на русском языке.

ЭНИ фольклорных фондов и архивов является динамической системой, так как задача представления в электронном виде всего фольклорного фонда БашГУ не достижима моментально. Отправным пунктом оформления ЭНИ фольклорных фондов и архивов БашГУ послужил фольклорный фонд кафедры башкирской литературы и фольклора факультета башкирской филологии и журналистики БашГУ.

На сегодняшний день фонд кафедры состоит из 145 томов машинописи и 53 томов рукописных текстов, которые и предполагается перевести в форму ЭНИ в первую очередь.

Записи представлены в текстовом формате, снабжены метаинформацией и ссылкой на источник (дело и листы в архиве). Метаинформация включает название текста, его жанровую принадлежность, место записи, год, язык текста, источник публикации, имена фольклористов и информантов.

Помимо самих текстов и необходимой для публикации фольклорных записей метаинформации ЭНИ оснащено специфическими для электронного издания сервисами:

1. Интерактивное оглавление, позволяющее выстраивать опубликованные тексты в удобном пользователю порядке, группируя их по заглавию, жанру, году или месту записи.
2. Приложенные к каждой публикации интерактивные карты с отмеченным на них местом записи текста. Осуществлено на основе сервиса Google Maps.

3. Автоматически составляемый указатель словоформ издания. Каждую представленную в указателе словоформу сопровождают данные о частотности и ссылка, по которой можно посмотреть все контексты употребления формы в ЭНИ. Со страницы с контекстами можно перейти непосредственно к тексту в издании, где искомая словоформа будет подсвечена. Имеется также и специальный словарь частотности словоформ издания. Указатель обновляется автоматически и не зависит от графика вывески новых текстов.

4. Автоматически генерируемая библиографическая информация. Ссылка для цитирования текстов издания в научных работах размещается внизу каждого документа под заголовком «Ссылка на эту публикацию». Ссылка соответствует ГОСТ 7.0.5–2008 «Система стандартов по информации, библиотечному и издательскому делу. Библиографическая ссылка. Общие требования и правила составления» и включает в себя информацию о заглавии текста, названии и годе издания, а также о дате публикации текста и дате обращения пользователя.

5. Экспорт в формат XML. Extensible Markup Language (XML) — рекомендуемый научным сообществом формат филологических электронных публикаций. Все помещённые в ЭНИ тексты доступны для пользователя в этом формате в соответствии со стандартами международного консорциума TEI: Text Encoding Initiative.

Оглавление включает в себя название текста (оно одновременно является ссылкой, по которой можно перейти к самой записи), жанровое определение, указание на место записи, сокращённое до района (полную информацию, включающую название населённого пункта можно посмотреть на странице записи в блоке метайнформации), год записи, язык текста и источник публикации. В ряде случаев название даётся издателем, в оглавлении это специально не отмечается, но на странице записи такие названия помещены в угловые скобки. Жанровые определения даются в сокращённом виде, полностью они представлены на странице записи.

По щелчку на заголовке столбца с названием текста, жанровым определением, местом записи, годом, языком или источником, можно изменить порядок расположения записей в оглавлении, выстроив его в прямом (один щелчок) или обратном (два щелчка) алфавитном или числовом порядке этого столбца.

В ЭНИ представлены тексты, записанные в Салаватском, Кугарчинском, Зианчуринском, Баймакском, Бурзянском районах Республики, а также в Оренбургской области. Районы работы фольклористов обозначены на карте на одной из интернет-страниц ресурса.

Созданная электронная платформа для публикации фольклорных текстов позволит в будущем как расширять и совершенствовать функциональность, так и проводить систематические исследования на основе оцифрованной коллекции.

Главным направлением, в котором будет развиваться сопровождающий издание инструментарий исследователя, — это разнообразные поисковые механизмы. Уже сейчас в коллекции можно искать словоформы, находить контексты их употребления и знакомиться с данными о частотности. Однако с точки зрения фольклориста — это «микроуровень». Гораздо более важными являются параметры более высокого, чем лингвистический, уровня, относящиеся к организации текста, его места в фольклорной системе.

Проще всего в этом смысле с повествовательными текстами. Для них давно разработаны и каталогизированы классифицирующие признаки. Для сказок это прежде всего персонажи с определённым социальным статусом и функциями (мужик, поп, солдат, мачеха, младший сын и т. д.), объекты, определяющие пространственные параметры повествования (река, гора, озеро, и т. д.), зооморфные персонажи (лиса, медведь, конь, и т.п.), а также специфические предметы, формирующие сказочные мотивы (веретено, меч, сун-дук и т.п.)

Эти признаки будут задействованы и в организации материалов ЭНИ.

В лирических жанрах такие параметры теряют свою релевантность. Однако во всех без исключения текстах так или иначе проявляют себя ключевые понятия языковой и фольклорной картины мира. Следующим исследовательским этапом после собственно публикаторского должно стать выявление этих понятий и системных связей между ними. Именно ЭНИ позволит в будущем на основе лексикостатистики, прикладных аспектов лингвистики текста, data mining и information extraction проводить исследования, которые будут не только классифицировать имеющиеся записи, но и давать ответы на насущные вопросы культурологии и фольклористики.