

УДК 004:9:002

## РУКОПИСНЫЕ И СТАРОПЕЧАТНЫЕ КИРИЛЛИЧЕСКИЕ КНИГИ В ИНТЕРНЕТЕ\*

*Д. А. Гагарина, С. И. Корниенко*

Статья посвящена анализу ресурсов Интернета, содержащих рукописные и старопечатные кириллические книги, их коллекции и каталоги. В ходе первичного анализа были отобраны 30 ресурсов, на основании которых рассматривались существующие способы организации каталогов и коллекций в Интернете, способы представления и визуализации изданий, возможности их чтения и исследования.

*Ключевые слова:* рукописные книги, старопечатные книги, кириллица, Интернет, распознавание текстов.

### *Введение*

Одной из составляющих процесса информатизации науки и образования является формирование современной информационной среды для исторической науки и гуманитарных наук вообще. Важнейшим компонентом этой среды становятся электронные версии исторических источников, в том числе размещенные в Интернете, позволяющие, с одной стороны, расширить доступ исследователей и учащихся к ним, а с другой – использовать современные методы их анализа.

Среди исторических источников достаточно существенное место принадлежит рукописным и старопечатным книгам. Они являются ценным историко-культурными памятниками и важным источником исторической информации. Формирование современной научной и образовательной среды гуманитарных наук невозможно без включения в нее и этой составляющей. Это подтверждается тем, что уже сегодня существует немало интернет-ресурсов, относящихся к рассматриваемой тематике, созданных в результате реализации различного рода исторических и информационных проектов.

В 2009 г. работа над одним из таких проектов началась и в Пермском государственном университете. Проект предусматривает создание программного комплекса для распознавания рукописных и старопечатных текстовых исторических источников. В перспективе предполагается разработка информационной системы на основе полнотекстовой базы этих источников.

В рамках подготовки и реализации проекта был проведен обзор ресурсов Интернета, связанных с рукописными и старопечатными кириллическими текстами. Его целью был количественный и качественный анализ представительства традиционной книжности в Интернете, выявление способов организации ресурсов и способов представления источников, необходимый для определения путей дальнейшего развития электронных публикаций в этой области.

### *Методика поиска и отбора ресурсов*

В основном поиск ресурсов осуществлялся с помощью поисковых систем. В качестве запросов использовались «рукописные и старопечатные книги», «рукописные кириллические книги» и др. Количественные результаты запросов к поисковым системам (табл. 1), несмотря на всю их формальность, показывают актуальность рассматриваемой темы.

Таблица 1

Результаты запросов к поисковым системам (сентябрь, 2009)

Запрос	Google	Яндекс
Рукописные и старопечатные книги	12800	28000
Рукописные и старопечатные тексты	4390	13000
Рукописные и старопечатные кириллические тексты	2110	1735
Рукописные и старопечатные кириллические книги	3870	1735
Рукописные кириллические книги	7780	29000

© Д. А. Гагарина, С. И. Корниенко, 2009

\* Исследование поддержано РФФИ (грант № 09-06-00254-а).

Рукописные кириллические тексты	5250	20000
Старопечатные кириллические тексты	2930	1503
Старопечатные кириллические книги	1200	2739
Старопечатные книги	27500	80000
Старопечатные тексты	10400	26000

Второй источник ресурсов – ссылки на родственных интернет-сайтах и коллекции тематических ссылок на таких крупных историко-ориентированных ресурсах, как «Электронная библиотека исторического факультета МГУ» [1], ХРОНОС [2] и др.

В ходе первичного анализа было отобрано 30 ресурсов [3–32], являющихся каталогами, коллекциями рукописных и старопечатных кириллических книг или ресурсами, посвященными одному такому изданию. Основной принцип первичного отбора ресурсов – их «полезность» с точки зрения изучения существующих направлений организации и представления рукописных и старопечатных источников в Интернете и связанных с этим проблем. Данная выборка вполне репрезентативно отражает состояние рассматриваемой области.

#### *Классификации ресурсов*

Все ресурсы, относящиеся к рассматриваемой тематике, могут быть разделены на те, которые содержат информацию об источниках, и те, которые содержат сами источники (их фрагменты).

Среди ресурсов первого типа можно выделить:

- каталоги;
- исследования;
- учебные материалы;
- ресурсы с описаниями источников;
- научно-популярные и справочно-энциклопедические ресурсы;
- ресурсы, посвященные методике описания источников и др.

Все эти ресурсы могут быть изначально электронными или электронными версиями печатных изданий (при этом среди ресурсов значительно преобладают последние). Электронные каталоги, будь они оригинальными или электронными версиями печатных, зачастую представлены просто электронным текстом и не содержат даже минимальной навигации внутри, не говоря уже о возможностях выборки и анализа.

Среди обозначенных категорий ресурсов в рамках данного обзора нам были интересны только электронные каталоги рукописных и старопечатных кириллических изданий, причем те, что изначально являются электронными и не ограничиваются простым текстом. Примерами таких каталогов являются «Книги кириллической печати XVI–XX вв. из собрания ГПИБ» [16], «Древлехранилище. Собрание старопечатных и рукописных книг» [8] и «Кириллические рукописные книги XVI–XX вв. в собраниях Челябинской области» [15].

Основной интерес представляют ресурсы, которые содержат непосредственно источники или их фрагменты. Каждое издание в таком ресурсе может быть представлено следующими составляющими:

- библиографическими описаниями;
- археографическими описаниями:
  - ✓ краткими;
  - ✓ подробными;
- изображениями (имиджами):
  - ✓ всех страниц;
  - ✓ нескольких страниц;
  - ✓ фрагментов страниц;
  - ✓ только переплета;
- текстами:
  - ✓ оригинальными шрифтами;
  - ✓ современными шрифтами;
- переводами;
- библиографическими списками с факсимиле издания или публикациями о нем.

В интернет-ресурсах, содержащих рассматриваемый тип изданий, каждый источник может быть представлен одним или несколькими названными составляющими в различных комбинациях. Наиболее информативным на сегодня является трехформатное представление источников (изображение – оригинальный текст – перевод), однако для рассматриваемого типа ресурсов такой вариант не характерен.

Покажем, как названные составляющие представлены в рассматриваемых интернет-ресурсах.

Полные археографические описания с росписью листов представлены в разделе «Рукописные книги» на сайте Национальной библиотеки Карелии [22], в электронном каталоге «Кириллические рукописные книги XVI–XX вв. в собраниях Челябинской области» [15] и в проекте «Древнерусские берестяные грамоты» [11], остальные ресурсы представлены краткими археографическими описаниями, библиографическими описаниями, свободным описанием изданий или разными типами описаний для разных книг коллекции. Треть ресурсов вообще не содержит описаний (ни библиографических, ни каких иных), порой вызывает затруднения даже определение хронологической принадлежности источника.

Треть ресурсов не содержит изображений изданий. Примерно пятая часть ресурсов содержит полные коллекции изображений страниц изданий. К ним, например, относятся «Дом Живоначальной Троицы» [7], «Соборник» [27] и др. [21; 22; 26; 30].

Изображения могут быть сделаны на основе оригинала или копии (факсимильного издания) и представлены следующими форматами: jpeg, gif, pdf, djvu и др. Явным «лидером» является jpeg. В ряде сайтов встречаются псевдоизображения, представленные в формате pdf, полученные с помощью соответствующих программ и близких к оригиналу шрифтов, примерами являются проекты «Библиотека святоотеческой литературы» [5], «Православная страница Николая и Елены Андрущенко» [20] и «Славянская Библия» [25].

Наличие распознанных текстов характерно примерно для половины из рассмотренных ресурсов. Впрочем, их большое количество в рассматриваемой выборке объясняется принципами отбора ресурсов для исследования. В целом же для ресурсов подобных типов наличие электронных текстов скорее не характерно.

Форматы, в которых представлен текст источников, варьируются, это могут быть html, pdf, doc, ttf и др. Особенности шрифтов и алфавитов рукописных и старопечатных кириллических изданий ограничивают возможности их представления традиционными способами. Для их корректного представления и визуализации требуется использование специальных шрифтов, а также специальных кодировок. Часть ресурсов, содержащих рукописные и старопечатные тексты, как раз посвящены описанию таких шрифтов и кодировок.

Количество специальных шрифтов (гарнитур), используемых для рукописных и старопечатных текстов, достаточно велико: «Novgorod» [11]; «Putiata»; комплект шрифтов «Menaion» [19]; «Irmologion»; «Ostrog»; «Zlatoust»; «Indycton»; «Bukvica» [13]; «Orthodox» [20]; «Triod» [28]; «SBibSlav» [25] и др. Разработаны кодировки для разных типов графики письма (устав, вязь и др.) и даже для буквиц (рис. 1). Коллекции шрифтов представлены на сайтах «Ирмологий» [13] и «Типикон» [28].



Рис. 1. Шрифт «Bukvica», изображение с сайта «Ирмологий» [13]

Наиболее известной и распространенной кодировкой, используемой для записи и визуализации рассматриваемой категории текстов, является hir. Ее подробное описание можно найти на сайтах «Акафистник» [4] и «Библиотека святоотеческой литературы» [5], а используется она в целом ряде проектов, например [18; 20; 28] и др.

Для ввода электронных рукописных и старопечатных текстов разрабатываются виртуальные текстовые редакторы [19; 25]. Необходимость использования таких редакторов объясняется особенностями шрифтов и алфавитов, а также трудностями автоматизации ввода текстов.

Частично «обойти» необходимость визуализации специальными шрифтами можно в случае использования формата pdf с текстом под изображением – довольно распространенного варианта представления текстовых изданий. Однако низкое качество автоматического распознавания для текстов рассматриваемого периода не способствует применению данной разновидности формата pdf.

Помимо попыток представить «оригинальный» текст источников, используются варианты с текстами в современной орфографии или переводами. В целом количество переводов рукописных и старопечатных кириллических текстов в Интернете достаточно велико.

Интересными являются параллельные представления текстов. Параллельность может использоваться для оригинала и перевода, для различных переводов, для изображения и соответствующего распознанного текста. Рассмотрим некоторые примеры.

В проекте «Электронные публикации Института русской литературы РАН» [32] используется двухформатное представление текста – оригинальный текст и его перевод, размещенные в два столбца. Небольшой размер берестяных грамот позволяет расположить на одном экране цветное изображение грамоты, черно-белую прорисовку, оригинальный текст, перевод, а также описание источника [11]. Особенностью такого источника, как «Слово о полку Игореве», является большое количество его переводов, ресурс «Параллельный корпус переводов „Слова о полку Игореве“» [17] позволяет выбрать из списка нужные переводы и выдать их построчно вместе с оригиналом или без него.

Удобным для читателя является представление, помимо самих источников, также библиографических списков, посвященных их изучению, как, например, в проекте «Библиотека Фронтистеса» [6] и на портале «Манускрипт» [19].

Еще одним основанием для классификации рассматриваемых ресурсов может являться владелец сайта. По этому основанию ресурсы делятся на:

- ресурсы Русской Православной Церкви [28];
- ресурсы научных учреждений и научных сообществ [11; 14; 19; 32];
- ресурсы учреждений хранения (музеев, архивов, библиотек) [3; 16; 22; 30];
- частные ресурсы.

Кроме того, существуют совместные ресурсы. Так, размещение наиболее крупной в Рунете коллекции рукописей осуществляется в рамках совместного проекта Свято-Троицкой Сергиевой Лавры и Российской Государственной Библиотеки [7].

Принципы, по которым источники организованы в коллекцию (или каталог), могут быть следующими:

- соответствие месту хранения;
- демонстрации возможностей кодировок, шрифтов и т.п.;
- проблемно-ориентированные коллекции;
- коллекции источников одного типа и др.

Как правило, тип коллекции коррелирует с владельцем. Для учреждений хранения характерно размещение в Интернете коллекций, соответствующих их фондам. Отследить принцип, по которому ресурсы собраны в коллекцию, иногда не представляется возможным. Особенно это характерно для частных ресурсов.

С владельцем также, как правило, связаны цели создания ресурсов:

- исследовательские;
- расширение доступа;
- демонстрация возможностей кодировок, шрифтов и т.п.

Количество источников в рамках одного интернет-ресурса может варьироваться от одного до нескольких тысяч. На некоторых ресурсах посчитать количество представленных изданий не представляется возможным. В электронном каталоге «Древлехранилище» [8] представлены 5500 изданий, лидером среди коллекций является проект «Дом Живоначальной Троицы» [7] с почти тремя тысячами рукописей. На сайте «Древнерусские берестяные грамоты» [11] представлены около тысячи берестяных грамот.

При анализе выбранных ресурсов можно заметить следующую закономерность: количество изданий в рамках одного проекта коррелирует с качеством их описаний и представлений – большие по объему коллекции сопровождаются качественными описаниями. Последнее, естественно, не относится к проектам, изначально посвященным одному изданию. Среди таковых – «Слово о полку Игореве» [26], «Параллельный корпус переводов...» [17], «Псалтырь с Часословцем» [21] и др.

Важным является вопрос организации коллекции или каталога. Это может быть:

- простой список изданий;
- список разделов, возможно с подразделами;
- каталог с возможностью выбора атрибутов;
- каталог с полнотекстовым поиском;
- информационная система с аналитическими возможностями.

В большинстве случаев ресурсы содержат список изданий, разделенных на разделы. Принципы организации в разделы – хронологический, проблемный, по месту хранения и др. Иногда поиск конкретного издания в таком ресурсе на практике оказывается затруднителен.

Удобным для пользователя и оптимальным для больших коллекций и каталогов является возможность поиска по атрибутам, что реализовано в ресурсах [8; 11; 15; 16; 19]. Например, в проекте «Книги кириллической печати XVI–XX вв. из собрания ГПИБ» реализован текстовый поиск по таким атрибутам, как автор, название, место издания, типография, год издания, имена, географические названия [16]. Для удобства поиска по атрибутам могут использоваться справочники – например, справочник городов и типографий в каталоге «Древлехранилище» [8] и справочники городов, раскопов, сохранностей и жанров в проекте «Древнерусские берестяные грамоты» [11]. Каталог «Кириллические рукописные книги XVI–XX вв. в собраниях Челябинской области» снабжен системой справочников-указателей: «Указатель имен», «Географический указатель», «Указатель инципитов» и «Хронологическая таблица» [15].

Ряд проектов предоставляет не просто изображения страниц источников и их тексты, но также целый набор дополнительных возможностей, делающих удобным их чтение в Интернете. Наиболее распространенные из них:

- «листанье» страниц;
- переход на любую из страниц;
- увеличение страниц;
- оглавления с возможностью перехода на нужные разделы;
- скачивание книги целиком;
- просмотр текста в различных форматах и др.

Некоторые из обозначенных возможностей реализованы в проектах «Остромирово евангелие и рукописная традиция новозаветных текстов» [3], «Библиотека святоотеческой литературы» [5], «Дом Живоначальной Троицы» [7], «Соборник» [27], «Цифровая библиотека „Книжные памятники Сибири“» [30] и др.

Некоторые из проектов, посвященных рукописным и старопечатным изданиям, помимо самих источников и средств их просмотра, содержат различные сведения, относящиеся к рассматриваемой теме.

Так, комплексный научно-популярный ресурс «Русское письмо» содержит сведения, относящиеся к истории развития письма, сравнительные и хронологические таблицы, таблицы, в которых исторические события на Руси синхронизированы с литературными; изложение снабжено гиперссылками на изображения соответствующих фрагментов изданий и их описания [24].

Электронный каталог «Кириллические рукописные книги XVI–XX вв. в собраниях Челябинской области» [15], помимо самого каталога, содержит обширные сведения по истории изучения рукописных книг Челябинской области, принципам описания изданий, значительную библиографию, а также подробные описания самого каталога, его структуры и алгоритма использования, принципов построения, что представляет ценность само по себе.

Помимо расширения доступа к источникам и удобной их визуализации, существующие ресурсы могут предоставлять возможность осуществления компьютеризированных исследований. Пожалуй, наиболее широко такие возможности представлены на портале «Манускрипт», содержащем систему выборки и запросов, позволяющих подготовить данные для лингвистических, палеографических и текстологических исследований, морфологический анализатор для автоматического

анализа и синтеза словоформ древнерусского языка и целый ряд других возможностей [19]. Так, например, критическое издание служебной минеи на май, подготовленное в рамках проекта «Манускрипт», предоставляет возможность познакомиться со структурой и составом рукописей, получить информацию о соответствиях между структурными и лингвистическими единицами рукописей, получить сведения о различиях между рукописями, получить материал для анализа соотношений греческого и славянского текста [33].

### **Выводы**

Анализ ресурсов Интернета, содержащих рукописные и старопечатные кириллические издания, их коллекции и каталоги, позволяет сделать следующие выводы.

1. Несмотря на актуальность темы, количество таких ресурсов невелико, что делает особенно актуальной дальнейшую работу по публикации в электронном формате рукописных и старопечатных памятников.

2. Существующие ресурсы разноплановы и отличаются по количеству изданий, способам их организации и визуализации, предоставляемым возможностям. Практически каждый сайт уникален. Нет сложившихся канонов публикации источников рассматриваемого типа.

3. Большинство изданий представлено в графическом формате.

4. Публикации не в оригинальном виде, а в современной орфографии или переводах осуществляются достаточно широко. Однако в таких публикациях исчезает облик источника, шрифтовые и лингвистические особенности, «дух языка». Он становится вполне современным документом, информационный потенциал источника не может быть раскрыт в полной мере при переводе.

5. Достаточно большая часть ресурсов не удовлетворяет сложившимся требованиям публикации электронных версий книг: могут отсутствовать библиографические данные о печатном издании, сведения об авторах и редакторах электронной версии.

6. Лишь единичные ресурсы содержат большие коллекции и каталоги книг. В большинстве случаев ресурсы содержат до 20–30 изданий.

7. Хотя существуют технологии кодирования и представления текстов рассматриваемой категории, а также разработаны специальные шрифты, массив книг, которые с помощью них визуализированы, остается небольшим.

Обозначенные результаты и трудности являются следствием ряда причин и особенностей.

С одной стороны, это особенности самих исторических источников – рассматриваемому типу ресурсов свойственно разнообразие источниковой базы, что вызывает трудности с единообразным описанием, хранением и представлением источников.

С другой, технической, стороны, на сегодня отсутствуют удовлетворительные результаты автоматического распознавания рукописных и старопечатных кириллических текстов. Проблема совершенствования технологии автоматического распознавания текстов стоит достаточно остро. Несмотря на то что еще в 1999 г. в рамках ассоциации «История и компьютер» такая проблема была обозначена, более того в статье Д. М. Володихина [34, 178–181] прозвучал довольно оптимистичский прогноз относительно возможности применения для распознавания средневековых кириллических текстов систем, основанных на технологии OCR, этого до сих пор не произошло. Основным методом создания полноценных электронных текстов рассматриваемых источников остается ручной посимвольный набор. Необходимость использования специальных виртуальных редакторов и шрифтов еще более увеличивает трудоемкость. Пожалуй, отсутствие эффективных систем распознавания рукописных и старопечатных текстов является основным препятствием к наращиванию массива электронных текстовых публикаций этих источников, развитию компьютеризированных методов их исследования. Указанное обуславливает целесообразность развития работ, направленных на создание систем распознавания, основанных на наиболее совершенных методах и технологиях, прежде всего методах искусственного интеллекта.

### **Список литературы и источников**

1. Электронная библиотека исторического факультета МГУ [электронный ресурс]. URL: <http://hist.msu.ru/ER/index.html>
2. ХРОНОС – всемирная история в Интернете [электронный ресурс]. URL: <http://hrono.ru/>
3. On-line выставка «Остромирово евангелие и рукописная традиция новозаветных текстов» [электронный ресурс]. URL: <http://www.nlr.ru/exib/Gospel/>
4. Акафистник [электронный ресурс]. URL: <http://akafistnik.ru/>

5. Библиотека святоотеческой литературы [электронный ресурс]. URL: <http://www.orthlib.ru/>
6. Библиотека Фронтистеса [электронный ресурс]. URL: <http://ksana-k.narod.ru/>
7. Дом Живоначальной Троицы, Официальный сайт Троице-Сергеевой Лавры. Рукописи [электронный ресурс]. URL: <http://stsl.ru/manuscripts/index.php>
8. Древлехранилище. Собрание старопечатных и рукописных книг [электронный ресурс]. URL: <http://virlib.eunnet.net/depository/?nch=2>
9. Древнерусская литература [электронный ресурс]. URL: <http://old-russian.narod.ru/index.htm>
10. Древнерусская литература. Антология [электронный ресурс]. URL: <http://old-rus.narod.ru/>
11. Древнерусские берестяные грамоты [электронный ресурс]. URL: <http://gramoty.ru/>
12. Изборник [электронный ресурс]. URL: <http://litopys.narod.ru/>
13. Ирмологий [электронный ресурс]. URL: <http://irmologion.ru/>
14. Исторические источники по Отечественной истории до начала XVIII в. на русском языке в Интернете [электронный ресурс]. URL: <http://www.hist.msu.ru/ER/Etext/PICT/feudal.htm>
15. Кириллические рукописные книги XVI–XX вв. в собраниях Челябинской области [электронный ресурс]. URL: <http://unilib.chel.su:81/root2/kp/cyrillic/predislovie.html>
16. Книги кириллической печати XVI–XX вв. из собрания ГПИБ [электронный ресурс]. URL: <http://www.shpl.ru/project/rarebooks/database/>
17. Параллельный корпус переводов «Слова о полку Игореве» [электронный ресурс]. URL: <http://nevmenandr.net/slovo/>
18. Печатный двор [электронный ресурс]. URL: <http://pechatnyj-dvor.su/>
19. Портал «Манускрипт» [электронный ресурс]. URL: <http://manuscripts.ru/>
20. Православная страница Николая и Елены Андрущенко [электронный ресурс]. URL: <http://www.orthonord.ru/>
21. Псалтырь с Часословцем [электронный ресурс]. URL: <http://www.nlr.ru/e-res/psaltyr/>
22. Рукописные книги на сайте Национальной библиотеки Карелии [электронный ресурс]. URL: <http://library.karelia.ru/cgi-bin/library/folder.cgi?id=116>
23. Рукописные памятники Древней Руси. Электронный архив [электронный ресурс]. URL: <http://www.lrc-lib.ru/>
24. Русское письмо [электронный ресурс]. URL: <http://character.webzone.ru/index.htm>
25. Славянская Библия [электронный ресурс]. URL: <http://www.ipmce.su/~igor/bible.html>
26. Слово о полку Игореве [электронный ресурс]. URL: <http://old-rus2.chat.ru/>
27. Сборник. Собрание книг кириллической печати [электронный ресурс]. URL: <http://sobornik.ru/index.htm>
28. Типикон. Страница Игумена Корнилия [электронный ресурс]. URL: <http://typikon.ru/>
29. Фонд знаменных песнопений [электронный ресурс]. URL: <http://znamen.ru/index.php>
30. Цифровая библиотека «Книжные памятники Сибири» («Рукописные и старопечатные книги из сибирских хранилищ») [электронный ресурс]. URL: <http://www.spsl.nsc.ru/rbook/>
31. Шедевры коллекций [электронный ресурс]. URL: <http://www.nlr.ru/fonds/best/index.html>
32. Электронные публикации Института русской литературы (Пушкинского Дома) РАН [электронный ресурс]. URL: <http://lib.pushkinskijdom.ru/>
33. Электронное критическое издание майской служебной минеи (вер. 1.0) [электронный ресурс]. URL: [http://docs.google.com/present/view?id=dgg9jbrr\\_150g97487df](http://docs.google.com/present/view?id=dgg9jbrr_150g97487df)
34. *Володихин Д. М.* Перспективы применения программ оптического распознавания в работе со средневековыми русскими кириллическими текстами (Из опыта применения компьютерных технологий в преподавании русской палеографии) // Инф. бюллетень Ассоциации «История и компьютер». № 24. М., 1999.

*Дата поступления рукописи в редакцию: 24.10.2009*