

УДК 002.2:004:091:069.5:930.1(47+57)(092)
ББК 78.501+78.558+63.1(2)6–81

ТЕХНОЛОГИЯ СОЗДАНИЯ ПОЛНОТЕКСТОВОЙ БАЗЫ ДАННЫХ «РУКОПИСИ ИЗ СОБРАНИЯ АКАДЕМИКА М. Н. ТИХОМИРОВА»

© А. В. Шабанов, 2014

*Государственная публичная научно-техническая библиотека
Сибирского отделения Российской академии наук
630200, г. Новосибирск, ул. Восход, 15*

Представлены итоги работы по созданию цифровых копий рукописей из собрания академика М. Н. Тихомирова.

Ключевые слова: русские рукописные книги, цифровые библиотеки, обработка изображений.

The results of work on creating digital copies of manuscripts from the collection of academician M. N. Tikhomirova are presented.

Key words: Russian manuscripts, digital libraries, image processing.

В настоящее время завершена многолетняя работа по оцифровке рукописных книг из собрания академика М. Н. Тихомирова – основы фонда отдела редких книг и рукописей ГПНТБ СО РАН. Выполнение этой задачи потребовало решения целого комплекса взаимосвязанных проблем, в числе которых: разработка концепции представления книжного памятника, выбор оборудования, первичная оцифровка и дальнейшая обработка изображений, обеспечение системы хранения данных и доступа к созданным базам данных (БД), а также цифровым библиотекам, разработка прикладного программного обеспечения, включение в состав электронной копии научного описания.

Далее мы изложим итоговые решения и результаты, подробно отраженные в работах В. Н. Алексеева, А. Ю. Бородихина, Е. И. Дергачевой-Скоп, А. В. Шабанова и др. [1–7].

О собрании академика М. Н. Тихомирова, которое он передал в 1965 г. Сибирскому отделению Академии наук СССР, писали, например, А. И. Рогов и Н. Н. Покровский [8]. Всего в нашем фонде хранится 649 рукописей этого собрания, которые послужили основой для дальнейшего развития отдела редких книг и рукописей ГПНТБ СО РАН. Так, за последующие почти 50 лет активной археографической работы на территории Сибири и Дальнего Востока фонд пополнился еще примерно 600 рукописями. Кроме этого, в составе подаренного собрания около 100 старопечатных книг, но о них в настоящей работе не рассказывается. Самостоятельно можно изучить цифровые копии трех изда-

ний Ивана Федорова¹ на интернет-сайте www.spsl.nsc.ru/rbook в соответствующем разделе.

Разработка концепции представления книжного памятника выполнена в рамках интеграционного проекта Президиума СО РАН № 7 «Принципы и технология электронного представления книжного памятника» (2009–2011 гг.) и в сжатом виде изложена в работе В. Н. Алексеева, Е. И. Дергачевой-Скоп и др. [5]. Отдельную трудоемкую часть при создании современной электронной копии составляет научное описание. Его принципы и структура разработаны в интеграционном проекте Президиума РАН № 25.7 «Тихомировское собрание рукописей ГПНТБ СО РАН: научное описание неизвестной в науке части собрания, оцифровка как обеспечение качественно нового доступа к рукописным подлинникам». Примеры создания электронных копий, включающие научные описания, даны, например, в работе А. В. Шабанова [4].

Выбор оборудования – наиболее ответственная часть всей работы, поскольку ошибки, допущенные на этом этапе, трудно исправить (необходимы большие затраты денежных средств и времени). Мы подробно изучили разные цифровые фотоаппараты, объективы к ним, системы освещения и небольшой ряд планетарных сканеров. В результате для проведения оцифровки редких книг и рукописей в ГПНТБ СО РАН были выбраны две установки:

¹ Апостол. Львов, 1574 г. Печ. – Ив. Федоров и П. Т. Мстиславец. Тих. 55-К, Новый завет с Псалтырью. Острог, 1580 г. Печ. – Ив. Федоров. Тих. 41-К, Библия. Острог, 1581 г. Печ. – Ив. Федоров. Тих. 22-К.

типа ПланСкан «Репро» фирмы ЭЛАР и на основе цифрового фотоаппарата Canon 500D с объективами Canon EF 24 mm f/1.4L USM и Canon EFS 18-55 mm и двух источников рассеянного света FalconEyes Softbox SBQ-7575. В рабочем помещении создан режим «темной комнаты» для стабилизации условий оцифровки.

Стоит отметить, что выбор оборудования зависит от состава коллекций фонда, а его цена колеблется от 50 тыс. до 10 млн руб. При этом обязательно выбирать самое дорогое, на рынке присутствует довольно широкий круг предложений.

Цель обработки изображения – получить высококачественную цифровую копию, адекватно отображающую первоисточник и пригодную для работы специалистов при отображении на экране монитора уровня Full HD. Создание архива изображений для последующего репринтного издания не является обязательным и обеспечивается только для небольшого круга уникальных источников. Обрабатывают изображения в программах FastStone Image Viewer (<http://www.faststone.org>) и Adobe Photoshop CS3 (<http://www.adobe.com>). Основные базовые операции: поворот и кадрирование первичного изображения, коррекция цвета и усиление резкости, изменение размеров в зависимости от требований к цифровой копии. Для ряда особо проблемных с точки зрения адекватного представления электронной версии на различных устройствах вывода используют более сложные алгоритмы обработки изображений [8]. Самостоятельно проанализировать полученные результаты можно с помощью данных, доступных по адресу <http://www.spsl.nsc.ru/rbook/examples/Tix-373/index.html>.

Следует особо отметить, что методы обработки изображений стремительно развиваются, значит вполне возможно возвращаться к архивам первичных изображений для создания цифровых копий более высокого уровня. Поэтому надежное сохранение исходных данных является (почти для всех электронных коллекций) обязательным. Мы используем 2-кратное дублирование данных в сетевом хранилище, частично на стримере, и дополнительно создаем еще одну копию на жестких дисках, хранящихся в отдельном месте и не подключенных постоянно к компьютеру (только по мере необходимости добавления информации винчестер временно подключается к системе).

Полнотекстовая БД «Рукописи из собрания академика М. Н. Тихомирова» генерируется на основе более широкой БД, создаваемой для ведения цифровой библиотеки «Книжные памятники Сибири» в рамках интегрированной расширяемой библиотечно-информационной системы (ИРБИС).

Изложим точную последовательность действий по формированию вышеуказанной БД, опираясь на книгу М. Н. Тихомирова «Описание Тихомиров-

ского собрания рукописей» (1968, 194 с.), где описаны первые 500 рукописей, сформировано базовое описание и интеграционный проект Президиума РАН № 25.7, в котором предварительно описаны остальные 149 рукописей. После этого следовало или уточнение сведений о первоисточнике согласно данным, накопленным за последние 40 лет, или новое описание с учетом развития этой области в последнее время.

Непосредственно для формирования полей БД используется АРМ «Каталогизатор». Заполняются поля на следующих вкладках:

- «Дублетность» – поля 200, 210 и 215 (заглавие, год издания, объем);
- «Расширенное» как повторы поля 300 (размер, количество листов, почерк, филигранны, переплет, записи, оформление, физическое состояние, ссылка на источник, откуда взято описание). Для отдельных рукописей (певческих) – нотация, распев;
- «Технология» – поле 951 (ссылки на цифровые копии в форматах PDF и HTML с JPEG);
- «Содержание» – поле 330 (оглавление как повторения этого поля);
- «Редкие» – необходимые поля.

Прикладное программное обеспечение разрабатывают для обеспечения требуемых пользователю функциональных возможностей цифровой копии с учетом включения в ее состав современного научного описания. Форматы данных изображений: TIFF, JPEG, RAW. Цифровые копии создаются в форматах HTML и PDF. Основной язык программирования – JavaScript. Он позволяет довольно просто создать базовые, основные, необходимые конечному пользователю функции – «листание» электронной версии, увеличение размера страницы и отсылка к описанию источника и дополнительной информации. Кроме того, современные ревизии JavaScript включают расширенные инструменты по обработке изображений и потенциально помогают создавать цифровые копии расширенного функционала.

Для каждой рукописи собрания формируются две цифровые копии – в форматах PDF и HTML с JPEG. Для создания PDF файла используется программа i2pdf (freeware) с изображениями формата JPEG размера 1–3 Мб. Итоговые файлы объемом от 10 Мб до 1 Гб позволяют работать с цифровой копией на мониторе как с почти полным аналогом печатного источника. При необходимости можно напечатать отдельные страницы до формата А4 высокого качества.

Вторая цифровая копия (HTML с JPEG) содержит все варианты изображений, полученных при обработке первичного архива форматов TIFF или RAW. Размеры JPEG файлов при этом меняются от 100 Кб до 10 Мб, в результате мы имеем

электронные версии, пригодные для самых разных категорий пользователей в рамках их полномочий в системе доступа к данным, используемой ГПНТБ СО РАН. Доступ к первичным архивам изображений не предоставляется. Фактически это цифровой аналог самого собрания фонда редких книг и рукописей и частично решает взаимосвязанные сложные проблемы сохранности и доступности. По мере развития методов обработки изображений и повышения мощности компьютеров часть электронных материалов используется для создания цифровых копий более высокого качества.

Подведем итоги. Всего оцифровано 415 рукописей. Для 310 из них на начало 2014 г. созданы электронные копии, для остальных 105 источников обработка изображений будет завершена к концу 2014 г. Для оцифровки 86 рукописей требуется более современное оборудование, так как это источники с мельчайшими деталями. Наконец, 148 рукописей нуждаются в реставрации различной сложности, перспективы их оцифровки пока неясны. В рамках настоящей работы привести полный перечень не представляется возможным. Наиболее примечательные источники даны в Приложении.

Общий объем полнотекстовой БД «Рукописи из собрания академика М. Н. Тихомирова» составляет около 560 Гб.

Значительная часть электронного собрания доступна без ограничений в Интернете www.spsl.nsc.ru/gbook (первый раздел). Полную версию можно получить в ГПНТБ СО РАН в рамках единой системы доступа к информации зарегистрированных пользователей.

Перспективы дальнейшего развития цифровой библиотеки отдела редких книг и рукописей ГПНТБ СО РАН связаны с созданием электронных версий старопечатных и рукописных книг региональных собраний, сформировавшихся в результате проведения в течение без малого 50 лет археографических экспедиций на территории Сибири и Дальнего Востока и генерации цифровых копий ряда источников из «живых» старообрядческих библиотек и частных коллекций. Отдельного внимания

заслуживает пополнение фонда научно-справочной литературы электронными версиями каталогов редких книг и рукописей, сводов филиграней и т. п.

Литература

1. Создание электронных версий старопечатных, рукописных и редких книг из сибирских хранилищ // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества / В. Н. Алексеев [и др.] : материалы Междунар. конф. «Крым 2001» (Судак). – М., 2001. – Т. 1. – С. 309–311.
2. Цифровая библиотека древнерусских книжных памятников Сибири в ГПНТБ СО РАН / В. Н. Алексеев, Е. И. Дергачева-Скоп, А. В. Шабанов, А. Ю. Бородихин // Современные информационные технологии и письменное наследие: от древних текстов к электронным библиотекам. Е1Manuscript-08 : материалы Междунар. науч. конф. (Казань, 26–30 авг. 2008 г.). – Казань, 2008. – С. 10–12.
3. Шабанов А. В. Факторы, влияющие на выбор технологии оцифровки русских старопечатных и рукописных книг // Библиосфера. – 2008. – № 4. – С. 46–48.
4. Шабанов А. В. Сравнение установок для оцифровки русских старопечатных и рукописных книг и методы обработки изображений // Библиосфера. – 2010. – № 2. – С. 30–32.
5. Цифровая версия Острожской Библии: к проблеме максимального функционального приближения электронного образа к «живому» книжному памятнику / В. Н. Алексеев, Е. И. Дергачева-Скоп, А. В. Шабанов, А. Ю. Бородихин // Остромирово евангелие и современные исследования. – СПб., 2010. – С. 84–93.
6. Бородихин А. Ю., Шабанов А. В. Цифровые копии учительного Евангелия ранней рукописной и печатной традиции // Библиосфера. – 2012. – Спецвып. – С. 94–96.
7. Шабанов А. В. Обработка изображений при создании цифровых копий рукописей с угасающим текстом // Труды ГПНТБ СО РАН. – Новосибирск, 2013. – Вып. 5 : Вклад ГПНТБ СО РАН в развитие отечественного библиотековедения, библиографоведения, книговедения и информатики. – С. 213–218.
8. Покровский Н. Н., Рогов А. И. Собрание рукописей академика М. Н. Тихомирова, переданное Сибирскому отделению АН СССР (г. Новосибирск) // Археографический ежегодник за 1965 год. Посвящен памяти академика М. Н. Тихомирова. – М. : Наука, 1966. – С. 162–172.

Приложение

Рукописные книги, включенные в состав цифровой библиотеки из собрания М. Н. Тихомирова (часть цифровых копий)

- № 1. Друцкое Евангелие. Пергаменная рукопись 1-й пол. XIV в.
- № 2. Соборное Уложение 1649 г. Рукопись 2-й пол. XVII в.
- № 6. Златоструй и Адриатис. Рукопись сер. XVI в.
- № 7. Слова Григория Богослова. Рукопись 1360-х гг.
- № 8. Слова Григория Богослова. Фрагмент пергаменной рукописи сер. XIV в.
- № 9. Лествица Иоанна Синайского. Рукопись 2-й пол. XVII в.
- № 11. Богородичник. Рукопись 1545 г.
- № 16. Евангелие-тетр. Рукопись конца XV в.

- № 20. Апостол. Рукопись 1-й пол. XVI в.
№ 22. Евангелие-тетр. Рукопись сер. XVI в.
№ 24. Евангелие-тетр. Рукопись сер. XVI в.
№ 26. Сказание о Тихвинской иконе Богородицы. Рукопись 2-й пол. XVIII в.
№ 29. Евангелие-апракос. Рукопись 1-й пол. XVI в.
№ 30. Евангелие-тетр. Рукопись сер. XVI в.
№ 36. Евангелие-тетр. Рукопись нач. XVI в.
№ 38. Триодь цветная. Рукопись нач. XVI в.
№ 40. Евангелие-апракос. Рукопись 1516 г.
№ 71. Азбуковник. Рукопись конца XVII в.
№ 72. Тропник. Рукопись XVII в., с добавлениями XIX в.
№ 87. Сборник Слов и Житий. Рукопись конца XVI – начала XVII в.
№ 95. Фрагмент Минеи служебной. Рукопись конца XIV – начала XV в.
№ 102. Апостол. Рукопись сер. XVI в.
№ 115. Сборник слов и житий, аскетический. Рукопись середины XV в.
№ 131. Апостол. Рукопись середины XVI в.
№ 132. Апостол. Рукопись 1647 г.
№ 150. Потребник. Рукопись начала XVII в.
№ 153. Жития Филиппа митрополита и Германа Соловецкого. Рукопись 2-й половины XVII в.
№ 164. Слово об иночестве. Рукопись XVII в.
№ 176. Апостол. Рукопись 1-й трети XVI в.
№ 191. Октоих крюковой. Рукопись 2-й половины XVIII в.
№ 280. Сборник Житий и Слов. Рукопись 1-й трети XVI в.
№ 295. Сборник с Житиями русских святых. Рукопись первой половины XVI в.
№ 296. Сборник Житий и Слов. Рукопись конца XV в.
№ 309. Сборник с Житиями святых. Рукопись первой четверти XVI в.
№ 351. Цветник со сказанием о граде Китеже. Конволют рукописей XVII–XVIII вв.
№ 406. Апостол. Рукопись конца XV в.
№ 492. Апостол. Фрагмент пергаменной рукописи XIV в.
№ 506. Триодный Торжественник. Рукопись XVIII в.
№ 520. Пролог (декабрь-февраль). Рукопись нач. XVI в.
№ 521. Златоуст. Рукопись конца XVI в.
№ 536. Кормчая (фрагмент). Рукопись первой четв. XV в.
№ 539. Кормчая. Рукопись середины XV в.
№ 554. Казанский летописец. Рукопись первой пол. XVIII в.
№ 581. Триодь постная и цветная на крюковых нотах. Рукопись XVII в.
№ 589. Сборник богослужебный на крюковых нотах. Рукопись XVII в.
№ 598. Сборник с Житием Корнилия. Рукопись-конволют XIX в.
№ 613. Пролог (март-август). Рукопись конца XV в.
№ 649. Портреты старообрядческих наставников и писателей, XVIII в.

Материал поступил в редакцию 18.02.2014 г.

Сведения об авторе: *Шабанов Андрей Васильевич – кандидат технических наук, старший научный сотрудник отдела редких книг и рукописей, тел.: (383) 266-10-91, e-mail: shabanov@spsl.nsc.ru*