

О. А. Юрченко

Научная библиотека Дальневосточного федерального университета

Решение проблем оцифровки библиотечного фонда: опыт Научной библиотеки Дальневосточного федерального университета

Рассмотрен опыт Отдела электронной библиотеки Научной библиотеки Дальневосточного федерального университета по оцифровке библиотечного фонда. Затронуты технические проблемы, возникающие в процессе оцифровки документов специалистами. Обоснован выбор графического формата TIFF в качестве основного для сканирования и PDF – основного для хранения пользовательской копии. Как альтернатива программному обеспечению сканера предложена бесплатная лицензионная программа IrfanView – графический редактор со встроенной функцией сканирования, что облегчает и ускоряет процесс. Для упрощения трудоёмкого процесса редактирования графических образов дополнительно к стандартным программам предложено использовать бесплатную лицензионную программу FastStone Image Viewer, которая также является графическим редактором. В программе FastStone Image Viewer имеются удобные функции пакетной обработки отредактированных графических образов, в частности преобразование в PDF с одновременным уменьшением веса конечного документа, что экономит объём электронного хранилища. Применение предложенных альтернативных программ, онлайн-ресурсов, а также способов, позволяющих устранять дефекты сканирования и избежать дефекты сканирования, значительно облегчает работу специалиста по оцифровке.

Ключевые слова: оцифровка документов, электронная копия, пользовательская копия, электронная библиотека ДВФУ, библиотечные фонды, обработка отсканированных документов.

Olga Yurchenko*Research Library of the Far Eastern Federal University, Vladivostok, Russia*

Solutions for library collection digitization: The experience of the Far Eastern Federal University Scientific Library

The experience of the E-library Department of the Far Eastern Federal University Scientific Library in digitizing the library collections is examined. The technological problems are discussed. The choice of TIFF graphic format for scanning and the PDF – for storing user copies is substantiated. Free license program IrfanView, a graph editor with a scanning embedded function which speeds up and smoothenes the process, is suggested as an alternative scanner software. To ease and speed up the labor-intensive process of graphic images editing to supplement standard software, the author suggests to use FastStone Image Viewer, a free license program. The FastStone Image Viewer offers the useful function of edited graphic images batch processing, in particular conversion to PDF with simultaneous document compression which saves the digital depository storage volume. The suggested alternative programs, online resources, and methods to eliminate scanning defects, as described by the author, would significantly lighten the work of digitization specialist

Keywords: document digitization, digital copy, user copy, Far Eastern Federal University E-library, library collections, scanned documents processing programs.

According to the "Recommendations on the digitization of materials from library funds" of the Expert council on digital copies, it is acceptable to create graphic images in the formats JPEG, TIFF, PNG. TIFF format is preferable when scanning books containing texts, signs and graphic images with a large depth of color. We used the free licensed program IrfanView, which has a convenient interface. It allows you to save the stability of the scanner settings and monitor them. In the research library of Far Eastern Federal University post-processing of scanned documents is performed in Picture Manager, Microsoft Paint and others. A free FastStone Image Viewer 5.9 program, which is designed for processing scanned images, is very convenient for use. Electronic copies of the document are stored in PDF format. This format is widespread, which completely eliminates the problem of format compatibility. The software that allows you to read a PDF document can be installed on any device, regardless of the operating system on it. When deciding on the choice of the storage format, its standardization by the organization ISO (International Organization for Standardization) was also of considerable im-

portance. This format already has the status of a standard for storing archival documents and for the exchange of information by companies in electronic form. Some libraries today use DjVu format for user copies, the main advantage of which is a very small amount of files, as well as the ability to organize a built-in interactive table of contents, which makes it possible to implement convenient navigation, saving disk space remains relevant. To accomplish this task and reduce the amount of the user copy, it is also advisable to use the FastStone Image Viewer program. It is enough to select the option "create a multi-page file", combining all the source files of the document into it and specify the desired quality in percent (the created file is automatically placed in the folder with the original images). As a result of using the program, the amount of the final file is significantly reduced. So, for example, for a rare book with color pages with a data volume with original images of 8.29 GB, the amount of the custom copy is 222 MB.

В Научной библиотеке Дальневосточного федерального университета (ДВФУ), как и в большинстве библиотек в последние годы, активно идёт процесс оцифровки собственного библиотечного фонда. Это обусловлено тем, что сканирование даёт более широкие возможности для предоставления информационного ресурса читателю и позволяет обеспечить лучшую сохранность оригинального документа [1].

Начало процессу оцифровки было положено около 11 лет назад научными библиотеками Дальневосточного государственного университета и Дальневосточного государственного технического университета, ныне входящих в состав ДВФУ. В настоящее время благодаря этому процессу НБ ДВФУ формирует электронную библиотеку собственной генерации.

Оцифровка бумажных изданий – сложная и трудоёмкая работа, поскольку у библиотек нет единой нормативно-методологической базы по выбору методов оцифровки и сканирующего оборудования [Там же]. На вебинаре «Создание качественного цифрового документа: опыт Национальной библиотеки Республики Карелия» сотрудники библиотек отметили отсутствие технической базы и специалистов по оцифровке в своих библиотеках [2]. Многогранные, как оказалось, проблемы приходится решать на практике.

В этой статье выделен ряд технических проблем, с которыми столкнулись сотрудники НБ ДВФУ, и предложены пути их решения, не требующие дополнительных финансовых затрат.

Выбор графического формата для сканирования. Согласно «Рекомендациям по оцифровке материалов из фондов библиотек» Экспертного совета по цифровым копиям [3] допустимо создавать графические образы в форма-

тах JPEG, TIFF, PNG. Ранее в НБ ДВФУ использовался формат JPEG, при сохранении в котором чертежей, текстовой и знаковой графики из-за сжатия происходят повреждения, потери.

Из опыта работы специалистов в области дизайна [4] известно, что формат TIFF более предпочтителен при сканировании книг, содержащих тексты, знаки и графические изображения с большой глубиной цвета. Формат TIFF широко применяется в полиграфии, входит во многие программы для обработки графических изображений, не имеет значительных потерь при обработке, не сжатый. Именно поэтому в НБ ДВФУ было принято решение сохранять графические образы в TIFF.

Нерезкие буквы в результате сканирования. Это результат работы сканера с установленным низким разрешением (количество точек на дюйм). Изменить ситуацию можно, увеличив разрешение в настройках сканера (оптимальное – 300 dpi). Следует отметить, что при этом увеличится и объём файла. Необходимо поддерживать качество сканирования. В качественно отсканированном изображении контуры букв чётко очерчены; буквы, включая самые мелкие элементы, равномерно плотно зачернены и хорошо просматриваются; околобуквенное пространство содержит пиксели только одного цвета – цвета фона.

К сожалению, сегодня нередко можно увидеть цифровые копии документов, у которых контуры букв нечёткие, плотность цвета букв разрежена. При увеличении видно, что пространство, окружающее буквы, содержит неоднородную смесь пикселей цвета как фона, так и букв. Изображение такого качества улучшить редактированием невозможно.

Нужное разрешение не сохраняется в настройках сканера, сбрасывается во время сканирования. Для облегчения работы со сканером производители нередко комплектуют оборудование программным обеспечением без возможности регулировать параметры сканирования. Как правило, по техническим условиям оборудования сканирование возможно и в интерфейсе другого программного обеспечения.

Существует множество программ, которые помогают более полно использовать возможности сканера, гибко настраивать процесс обработки изображений с регулировкой экспозиции, цветокоррекции, глубины цвета, формата сохраняемых файлов. Это такие платные программы, как *VueScan*, *ABBYY Scan Station* [5].

Столкнувшись с необходимостью каждый раз устанавливать параметры разрешения в настройках сканера, сотрудники библиотеки в качестве решения использовали бесплатную лицензионную программу *IrfanView* [6], имеющую удобный незагруженный интерфейс. В отличие от комплектного

программного обеспечения эта программа даёт возможность сохранить устойчивость настроек сканера и контролировать их.

Выбор программного обеспечения для постобработки отсканированных документов. Очень часто отсканированные документы нуждаются в последующей обработке. Существует ряд программ, позволяющих это сделать. Самая известная из них – *Photoshop*. Это очень мощный графический редактор, содержащий множество инструментов для профессиональной обработки графики и предоставляющий возможность и редактирования PDF-файлов, и пакетной обработки изображений, в том числе с целью уменьшения веса [7, 8].

Существуют и другие графические редакторы, при помощи которых можно редактировать сканированный документ. В Научной библиотеке ДВФУ постобработка отсканированных документов выполняется в программах *Picture Manager*, *Microsoft Paint* и др.

Очень удобна для использования бесплатная программа *FastStone Image Viewer 5.9*, предназначенная для обработки отсканированных изображений. Она даёт возможность выравнять наклон строк, обрезать страницы в один размер, сжимать нестандартные страницы до размера стандартных, а также осуществлять пакетную обработку полученных изображений, а именно перевод в PDF с одновременным сжатием. В результате получаются файлы небольшого объёма с чётким изображением букв, хорошо читаемые.

После сканирования на страницах появляются цветные дефекты. Чтобы цифровая копия отражала все особенности оригинальных документов, многие из которых содержат цветные иллюстрации и фоновые рисунки, а цвет страниц порой говорит о возрасте документа не хуже года его издания, принято выбирать цветной тип изображения в настройках сканирования.

Если в результате сканирования на странице образуется тень, потом она может окраситься цветом (даже если цвет в небольшом количестве присутствует на странице). Также на полученных в результате оцифровки изображениях могут быть резко выделяющиеся цветные точки и пятна – как ранее присутствовавшие на исходном материале, так и образовавшиеся в результате цифровой обработки. Нередко на отдельных изображениях появляется голубоватая подсветка букв.

Решить эту проблему можно с помощью функциональной бесплатной программы *FastStone Image Viewer* [6], в которую встроены эскизовый файловый менеджер и база данных, которая может быть опционально использована в качестве менеджера изображений и видео. Для устранения дефектов достаточно применить функцию «Коррекция цвета» и снизить до минимума параметр насыщенности.

Размер оригинала значительно превышает размер рабочей поверхности сканера. В этом случае возможно использование многофункционального, но в то же время простого в обращении графического редактора *Microsoft Paint*. Именно он поможет соединить документ, отсканированный частями.

Для этого в файл с одним из фрагментов, увеличенным до нужных размеров, необходимо вставить изображения из остальных фрагментов, правильно совместив все детали изображения. Увеличение размера цифровой копии страницы можно использовать также в тех случаях, когда не хватает размера полей: например, в типографии слишком много обрезали или имеются склеенные внутренними полями страницы, или после выравнивания наклона строк критично уменьшились размеры полей страниц.

Выбор формата пользовательской копии. В Научной библиотеке ДВФУ, как и во многих других библиотеках, электронные копии документа хранятся в широко распространённом формате PDF. Это полностью устраняет проблему совместимости форматов. Программное обеспечение, позволяющее прочитать PDF-документ, может быть установлено на любое устройство вне зависимости от используемой на нём операционной системы.

При принятии решения о выборе формата хранения немалое значение имела и его стандартизация *ISO (International Organization for Standardization)*. Этот формат уже имеет статус стандарта для хранения архивных документов и для обмена компаниями информацией в электронном виде.

Для пользовательских копий некоторые библиотеки сегодня используют формат DjVu, главным достоинством которого является очень маленький объём файлов, а также возможность организации встроенного интерактивного оглавления, что позволяет реализовать удобную навигацию. К сожалению, при очевидных преимуществах этот формат предназначен для хранения отсканированных копий документов, а в НБ ДВФУ оцифровка оригиналов не является единственным способом пополнения фонда электронной библиотеки собственной генерации. Кроме того, формат DjVu недостаточно распространён, а значит на подавляющем большинстве компьютеров отсутствует программное обеспечение для просмотра таких электронных документов.

Слишком большой объём пользовательской копии. Несмотря на стремительно развивающиеся облачные технологии, экономия дискового пространства остаётся актуальной. Для выполнения этой задачи и снижения объёма пользовательской копии также целесообразно использовать программу *FastStone Image Viewer*. Достаточно выбрать опцию «Создать многостраничный файл», объединив в него все исходные файлы документа, и указать желаемое качество в процентах (созданный файл автоматически размещается в папку с исходными изображениями).

В результате использования этой программы объём конечного файла значительно сокращается. Так, например, для редкой книги с цветными страницами при объёме данных с исходными изображениями 8,29 Гб объём пользовательской копии составляет 222 Мб.

Необходимо вырезать статью из электронной версии сборника / журнала. Если электронная версия сборника представлена в формате PDF, то для решения этой задачи подойдёт интернет-ресурс «Онлайн-инструменты для любителей PDF» (<http://www.ilovepdf.com/ru>), который содержит бесплатные онлайн-инструменты для объединения, разделения, сжатия файлов PDF, а также для преобразования файлов других форматов в PDF [9]. Этот ресурс имеет дружелюбный интерфейс, и работа с ним не вызывает сложностей.

Если электронная версия сборника представлена в формате DjVu, то необходимо использовать официальную бесплатную программу *DjVu Solo* [10] для редактирования, чтения и создания файлов в формате DjVu, а также интернет-ресурс конвертер «*DjVu to PDF*» (<http://djvu2pdf.com/>) [11]. На этом ресурсе за один раз обрабатывается до 20 страниц, поэтому следует предварительно подготовить нужную статью в исходном формате сборника DjVu и затем перевести в PDF.

Для изготовления статьи в DjVu в программе *DjVu Solo* создаётся новая книга и в неё загружается любой скан, который впоследствии будет удалён. Непременное условие: имя файла должно содержать 4 цифры перед расширением, например 1234.tif. После сохранения новой книги необходимо открыть в программе *DjVu Solo* исходный сборник, скопировать нужные страницы и вставить их в созданную книгу. Затем следует удалить ненужный первый файл и сохранить книгу. Так создаётся нужная статья в формате DjVu, которая переводится в PDF с помощью конвертера «*DjVu to PDF*».

Образование глубоких теней в месте сшива страниц. Если оцифровка выполняется на планшетных сканерах постранично или разворотом без разбора книги на листы, то на скан-копии в местах, где между рабочей поверхностью сканера и поверхностью страниц остаётся воздушное пространство, образуются глубокие тени чёрного цвета, переходящие на текст, а также искривляются строки. Как правило, это касается книг, имеющих узкие средние поля и склеенные переплёты.

Конечно, таких проблем можно избежать, если использовать специальные угловые книжные сканеры, в которых рабочая поверхность из стекла размещена не по центру, а продлена к переднему краю, или сканеры с камерой с V-образной колыбелью и с V-образным прижимным стеклом для книги.

Но даже при использовании обычного планшетного сканера можно предотвратить образование глубоких теней. Рабочие поверхности обычных сканеров имеют размеры, позволяющие наложить книгу только так, что каретка сканера будет последовательно проходить сначала первую страницу, затем вторую, при этом тень образуется только на второй странице. Таким образом, при постраничном сканировании, чтобы не образовывалась тень на странице, необходимо укладывать и переворачивать книгу так, чтобы нужная страница всегда была расположена первой по ходу движения каретки сканера.

Если книга сканируется разворотом, её надо укладывать так, чтобы каретка сканера одновременно проходила вдоль обеих страниц, т.е. от низа до верха страниц. Следует понимать, оцифровка книг разворотом возможна на сканерах с размером рабочей поверхности не менее А3.

Как уже было сказано, процесс оцифровки бумажных изданий – это сложная и трудоёмкая работа, так как электронные копии должны обладать свойствами, исключающими потерю информации и не ухудшающими читаемость документа по сравнению с бумажным оригиналом. Задачи, с которыми ежедневно сталкиваются специалисты библиотек в процессе оцифровки, разнообразны и не всегда имеют чёткие алгоритмы решения. Для их выполнения необходимы творческий подход и большое желание сделать свою работу лучше.

СПИСОК ИСТОЧНИКОВ

1. **Айгистов Р. А.** Оцифровка архивных и библиотечных документов / Р. А. Айгистов // Библиография. – 2015. – № 2. – С. 3–11.

Aigistov R. A. Otsifrovka arkhivnykh i biblioteknykh dokumentov / R. A. Aigistov // Bibliografiya. – 2015. – № 2. – S. 3–11.

2. **Третьякова С. В.** Создание цифрового документа: итоги вебинара для библиотекарей Карелии / С. В. Третьякова, Л. А. Карелина // Науч. и техн. б-ки. – 2016. – № 3. – С. 30–35.

Tret'yakova S. V. Sozdanie tsifrovogo dokumenta: itogi vebinaru dlya bibliotekarei Karelii / S. V. Tret'yakova, L. A. Karelina // Nauch. i tekhn. b-ki. – 2016. – № 3. – S. 30–35.

3. **Рекомендации** по оцифровке материалов из фондов библиотек [Электронный ресурс] / Экспертный совет по цифровым копиям. – Утв. 23.10.2013. – Электрон. текстовые дан. – Режим доступа: http://www.nlr.ru/pro/inv/digit_copy/recom.pdf.

Rekomendatsii po otsifrovke materialov iz fondov bibliotek [Elektronnyi resurs] / Ekspertnyi sovet po tsifrovym kopiyam. – Utv. 23.10.2013.

4. **Форматы** графических файлов [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.jetcom.ru/production/file-formats/>.

Formaty graficheskikh failov [Elektronnyi resurs].

5. **Популярные** программы Adguard Premium [Электронный ресурс] // Сайт компании Hamrick Software. – Электрон. дан. – Режим доступа: <http://vuescanpro.ru/>.

Populyarnye programmy Adguard Premium [Elektronnyi resurs] // Sait kompanii Hamrick Software.

6. **Официальный** сайт компании IrfanView [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.irfanview.com/>.

Ofitsial'nyi sait kompanii IrfanView [Elektronnyi resurs].

7. **Сидаш Д.** Как изменить размер сразу нескольких картинок (фотографий) [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://sidash.ru/kak-izmenit-razmer-srazu-neskolkih-kartinok-fotografii/>.

Sidash D. Kak izmenit' razmer srazu neskol'kikh kartinok (fotografii) [Elektronnyi resurs].

8. **Несколько** слов о сканировании и качестве [Электронный ресурс] // КомпьюАрт. – 2001. – Вып. 7. – Режим доступа: <http://www.i-type.ru/scanabout.html>.

Neskol'ko slov o skanirovanii i kachestve [Elektronnyi resurs] // Komp'yut. – 2001. – Вып. 7.

9. **Онлайн-инструменты** для любителей PDF [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.ilovepdf.com/ru>.

Onlain-instrumenty dlya lyubitelei PDF [Elektronnyi resurs]. – Elektron. dan.

10. **Официальный** сайт Djvu Solo [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://djvu-solo.ru/>.

Ofitsial'nyi sait Djvu Solo [Elektronnyi resurs].

11. **Конвертер** «DjVu to PDF» [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://djvu2pdf.com/>.

Konverter «DjVu to PDF» [Elektronnyi resurs].

Olga Yurchenko, Leading Librarian, Sector for Digital Technologies and Database Administration, Digital Library Department, Research Library, Far Eastern Federal University;

iurchenko.oa@dvfu.ru

65-B Aleutskaya st., 690090, Vladivostok, Russia