

Українська Православна Церква

**Т Р У Д И
Київської
Духовної
Академії**

№ 13



Київ
2010

Протоиерей Димитрий Юревич

**Преимущества использования
свободного программного
обеспечения при создании
электронных версий
дореволюционных книг (на примере
журнала «Христианское чтение»)¹**

Для чего нужны архивы отсканированных книг?

Интерес к технологиям сканирования и создания электронных книг в православных учебных заведениях вполне объясним. Гонения на Церковь в XX в. сопровождались практически полным уничтожением христианских библиотек и богословской литературы. Поэтому в начале 1990-х гг., когда церковь получила свободу действий и начала возрождаться, мы стали свидетелями книгоиздательского бума в православной среде. Наиболее интересные и значимые для православного богословия и аскетической традиции дореволюционные издания переиздавались репринтным способом. Это позволило в целом снять напряжение и удовлетворить потребность верующих в максимально востребованной литературе.

В 1990-е гг. проблема восстановления церковного книжного наследия не была решена до конца. Огромный пласт изданий по различным разделам богословия, библеистики, истории Церкви, церковно-практическим дисциплинам не был переиздан. Эти научные монографии и статьи представляют сегодня интерес — но немалый! — лишь для довольно узкого круга читателей, прежде всего — из среды преподавателей и студентов богословских учебных заведений. К числу неизданных трудов относятся, в первую очередь, дореволюционные журналы, которые издавались на протяжении многих лет (например, журнал Санкт-Петербургской духовной академии «Христианское чтение», который выходил до революции почти 100 лет — с 1821 по 1917 гг.). Переиздавать их репринтным способом небольшим тиражом было экономически неоправданно. При этом количество сохранив-

¹ Доклад на II Международной научно-практической конференции «Духовное и светское образование: история взаимоотношений — современность — перспективы» (Киевская духовная академия, 18 октября 2010 г.).

шихся копий дореволюционных изданий исчислялось единицами, и доступ к ним даже для заинтересованных специалистов был затруднен².

Поэтому, когда вначале 2000-х гг. появились доступные по стоимости сканеры, позволяющие получать изображения хорошего качества за приемлемое время, в православной среде начался труд по сканированию дореволюционной богословской научной литературы. Первый опыт представляли собой электронные издания журнала «Путь» и «Православной богословской энциклопедии», предпринятые в те годы Синодальной библиотекой Московского Патриархата имени Святейшего Патриарха Алексия II (директор — протоиерей Борис Даниленко). Самыми ранними, связанными с православными богословскими вузами, были проекты Перервинской православной духовной семинарии³ (руководитель — Сергей Архипов) и Санкт-Петербургского издательства «Аксион эстин»⁴ (директор — А. В. Егоров). Результаты их деятельности стали доступны публике в 2004 г. Год спустя, при моем непосредственном участии, подобный проект был организован при кафедре библеистики Московской православной духовной академии, в результате чего было издано 5 компакт-дисков с книгами. Часть из них была размещена также на сайте кафедры⁵.

В настоящее время потребность в создании отдела сканирования имеется практически в каждом православном учебном заведении. Количество неоцифрованной богословской литературы, включая оригинальные местные издания, до сих пор достаточно велико. Однако, организация постоянно продолжающегося процесса сканирования книг с их последующей обработкой и созданием электронного каталога имеет свои важные особенности и является более трудоемкой, по сравнению с нерегулярным сканированием одиночных изданий, которое предпринимают энтузиасты-одиночки.

В этом докладе мне хочется поделиться опытом работы отдела сканирования при Санкт-Петербургской православной духовной академии, созданного два года назад с целью оцифровки журнала «Христианское

²Как рассказал на пленарном заседании II Международной научно-практической конференции «Духовное и светское образование: история взаимоотношений — современность — перспективы» (проходила 18 октября 2010 г.) ректор Киевской духовной академии архиепископ Бориспольский Антоний (Паканич), при подготовке электронной версии журналов «Труды Киевской духовной академии» (1860–1879 гг.) поиск недостающих номеров журнала велся при содействии генерального директора Национальной исторической библиотеки А. В. Скорохватовой в разных библиотеках Киева. Чтобы собрать полный комплект «Трудов» за указанные годы, был задействован механизм межбиблиотечного абонементов.

³Часть книг размещена на их сайте <http://www.ppsds.ru>

⁴Ряд книг можно найти на их сайте <http://axion.org.ru>

⁵Это была одна из причин, по которой сайт кафедры библеистики МПДА (<http://bible-mds.ru>) стал победителем конкурса православных сайтов «Мрежа-2006» в номинации «Наука, культура, искусство, образование» (подробнее см.: <http://bible-mds.ru/events/2007-02-01/ev20070201-victory.html>).

чение»⁶. Одной из главных целей доклада будет изложить технологию сканирования с минимально возможными материальными затратами и результатом, отвечающим критериям кроссплатформенности и удобства для конечного пользователя.

Свободное программное обеспечение

При организации проекта сканирования принципиально важными требованиями для любого, не только православного, учебного заведения являются минимизация затрат и обеспечение эффективности работы сотрудников, не имеющих специальной компьютерной квалификации. Опыт работы отдела сканирования при Санкт-Петербургской православной духовной академии показывает, что полностью удовлетворить этим требованиям позволяет использование *свободного программного обеспечения* (Free Software) (далее — ПО). Эти программные продукты до сих пор еще не получили должного распространения в православных вузах. Здесь они могли бы стать полезны и эффективны не только в деле сканирования, но и для организации рабочих мест сотрудников и студентов.

Создание свободного ПО неразрывно связано с академической средой и ее научными интересами. К началу 1980-х гг. в сфере развития компьютерной техники сложилась крайне неприятная ситуация: всё новые компьютеры продавались только с платным программным обеспечением. Университеты были вынуждены приобретать такие компьютеры и платить за программы. Также не существовало возможности бесплатного распространения ПО и его модернизации. Даже в тех случаях, когда университеты имели возможность получить исходные коды программ и внести туда необходимые для себя изменения, они не могли распространять эти новые версии за пределами университета. Это сдерживало научные исследования и было неудобно в учебном процессе.

В январе 1984 года один из выдающихся американских программистов, а ныне еще и известный общественный деятель — Ричард Столлмен (Richard Stallman), выступил с инициативой создания операционной системы и комплекса программ, которые не только могли бы распространяться бесплатно, но и позволяли бы их пользователю вносить изменения в программный код⁷. В результате при последующих разработках ПО был сделан акцент на *свободу пользователя* при работе с компьютером. Четыре принципа свободы ПО были сформулированы следующим образом:

* пользователь имеет право запускать программу с любой, нужной ему целью;

⁶К настоящему моменту подготовлены и выпущены два выпуска DVD-дисков с электронным архивом журнала «Христианское чтение»: первый — в 2009 году (архив за 1893–1917 гг.), второй — в этом году (с 1868 по 1892 гг.). Сетевую версию можно найти по адресу <http://spbpa.ru>

⁷Столлмен Р. Проект GNU / Пер. с англ. С. Коропа // Сайт Фонда свободного ПО. URL: <http://www.gnu.org/gnu/thewnuproject.ru.html> (дата обращения: 15.10.2010).

* пользователь может модифицировать программу под свои нужды (для этого он должен иметь доступ к ее исходному коду);

* пользователь имеет право распространять исходную программу как бесплатно, так и за деньги;

* пользователь имеет право распространять свои модифицированные (улучшенные) версии программы⁸.

Изложенные идеи распространения ПО очень близки к христианскому принципу *даром получили — даром отдавайте* (Мф. 10, 8). Пользователь свободного ПО *сразу* (и, как правило — бесплатно) получает мощный инструментарий, результат труда многих людей — работоспособный программный комплекс, который может использовать в своей работе. Если возникает необходимость, пользователь имеет право *модифицировать* исходную программу для решения своих конкретных задач. В результате, вместо написания программы с нуля, что зачастую физически невозможно, он получает возможность путем некоторых доработок получить нужный продукт. Благодарной платой является *обязанность* для пользователя поделиться с компьютерным сообществом модифицированной версией программы и ее исходными кодами. Чтобы реализовать этот механизм был сформулирован принцип «copyleft» («авторское лево»), в противоположность понятию «copyright» («авторское право»). По словам Р. Столлмена, «copyleft использует систему авторского права, но в целях, противоположных ее обычному использованию: из способа приватизации программ оно превращается в средство защиты их свободы»⁹.

Проект, развивавшийся по этим принципам, получил название GNU (аббревиатура от Generally Not Unix), что в переводе означает «антилопа гну». Благодаря усилиям Р. Столлмена и ряда присоединившихся к нему программистов был создан Фонд свободного программного обеспечения. В сотрудничестве с рядом крупных корпораций Фонду удалось стимулировать создание новой операционной системы GNU. В качестве *ядра* этой системы было использовано ядро, которое начал разрабатывать в 1991 году финский студент Линус Торвальдс (Linus Torvalds) и которое, поэтому, получило название Linux (сейчас в его разработке участвуют сотни программистов, а Л. Торвальдс является координатором проекта). Именно этот компонент обеспечивал взаимодействие прикладных программ с конкретным оборудованием.

В настоящее время объединенный программный комплекс известен под названием GNU/Linux (более распространенное краткое, но неверное название — Linux) и существует в виде различных наборов программ, называемых *дистрибутивами*. В работе над дальнейшей разработкой свободной операционной системы GNU/Linux принимают

⁸ Столлмен Р. Проект GNU...

⁹ Там же.

участие тысячи программистов по всему миру, нередко получая деньги на это со стороны крупных корпораций (таких как Red Hat, Novell, IBM, занимающихся продвижением свободного софта, либо продажей оборудования). Идеи «копилефта» были юридически оформлены в виде текста лицензии GPL¹⁰, под которой сегодня выпускается подавляющее большинство свободных программ.

Технология создания архива отсканированных книг

Использование операционной системы GNU/Linux при организации проекта сканирования дает учебному заведению, прежде всего, немалую экономию средств. Ведь себестоимость необходимых программ, например в случае использования альтернативной и более распространенной в России операционной системы MS Windows, сопоставима со стоимостью компьютерного оборудования¹¹. По этой причине в качестве операционной системы в проекте сканирования Санкт-Петербургской православной духовной академии был выбран дистрибутив *Debian*¹² GNU/Linux (содержащий более 25 тысяч свободно распространяемых программ и модулей) и графическая среда KDE 3.5.

Технология создания электронного архива дореволюционных журналов логически может быть разделена на несколько этапов:

- * организация процесса сканирования;
- * обработка полученных изображений;
- * создание для каждой статьи электронного файла, включающего новый титульный лист с ее подробным библиографическим описанием;
- * создание электронного каталога, содержащего библиографические описания и ссылки на соответствующие файлы статей — как по номерам, так и по именам авторов.

На каждом этапе создания электронного архива имеются свои нюансы, относящиеся к работе с изображениями и использованию свободного ПО.

Организация процесса сканирования

Существуют различные способы получения изображений бумажных книг¹³. Предпочтительным вариантом является использование специальных сканирующих аппаратов. Они представляют собой авто-

¹⁰ GPL (сокр. от англ. «General Public License») — «Универсальная общественная лицензия GNU».

¹¹ Я сознательно уклоняюсь от анализа конкретных цен на несвободное ПО, поскольку они могут меняться в зависимости от страны и быть снижены в силу каких-либо акций компаний-производителей. В настоящее время набор несвободных программ, необходимых для организации процесса сканирования и обработки изображений (включая операционную систему), имеет тот же порядок стоимости, что и оборудование.

¹² Подробнее см. официальный сайт: <http://debian.org>

¹³ См., напр.: Как создать электронную книгу // Сайт «Atiz.ru». URL: http://www.atiz.ru/page.jsp?pk=node_1195748264247 (дата обращения: 15.10.2010).

матризованный комплекс, включающий два фотоаппарата для фотографирования страниц книги, механизм для перелистывания страниц и ПО для обработки. Мне довелось наблюдать за работой таких автоматов в Президентской библиотеке им. Б. Н. Ельцина в Санкт-Петербурге. По причине высокой себестоимости они недоступны для приобретения небольшому учебному заведению. Поэтому основным инструментом сканирования разумно считать планшетный сканер. Мы используем Epson Perfection (вначале это была модель 1270, сейчас — модель v10). Сканирование осуществляется программой xsane¹⁴ с разрешением 400 точек на дюйм (dpi; хотя приемлемым считается и 300 dpi), в черно-белом режиме, графический формат — tiff. Перед началом сканирования конкретной книги производится несколько пробных проходов для выбора вручную оптимальных настроек сканирования.

Процесс инсталляции конкретной модели сканера в GNU/Linux и сканирование с помощью xsane особых сложностей не вызывают (они были подробно описаны мной в свое время на сайте кафедры библеистики Московской православной духовной академии (далее — МПДА) и все желающие могут с ними там ознакомиться¹⁵).

Обработка отсканированных изображений

При обработке отсканированных изображений для автоматизации удаления черных полей и выравнивания по вертикали страницы текста используется программа ScanKromsator версии 5.6A, написанная русским программистом под никнеймом Volega¹⁶. Это — единственная программ из используемых нами, которая не входит в семейство свободных программ. ScanKromsator — бесплатно распространяемая программа под ОС Windows, которая без проблем запускается с помощью эмулятора

¹⁴ Официальный сайт: <http://xsane.org>. Эта программа представляет собой графический фронтэнд к консольной программе сканирования sane (официальный сайт: <http://www.sane-project.org>).

¹⁵ Юревич Д., свящ. Установка и настройка сканера под GNU/Linux: практическое руководство // Сайт кафедры библеистики МПДА. URL: <http://bible-mda.ru/soft/scanning/scanner-linux.html> (дата обращения 15.10.2010); Юревич Д., свящ. Сканирование изображений и документов в GNU/Linux на примере Epson Perfection 1270: практическое руководство // Сайт кафедры библеистики МПДА. URL: <http://bible-mda.ru/soft/scanning/scanning-linux.html> (дата обращения 15.10.2010).

¹⁶ На данный момент, насколько мне известно, у программы нет официального сайта. Версия 5.92 доступна для загрузки, например, отсюда: http://www.zodchii.ws/downloads/utills/scan_kromsator_v5_92_full.zip. Руководства можно найти здесь: Melirius. Использование ScanKromsator'a v5.91 // Сайт «Создание книг в электронном виде из бумажных книг (в формате DjVu)». URL: http://www.djvu-soft.narod.ru/kromsator/sk_5_91_melirius.htm (дата доступа: 15.10.2010); Monday2000. СканКромсатор v5.6A: пособие по программе // Сайт «Создание книг в электронном виде из бумажных книг (в формате DjVu)». URL: <http://www.djvu-soft.narod.ru/kromsator/> (дата доступа: 15.10.2010); Vadimir T. T. Создание электронных книг в формате DjVu // Сайт «Web-медицина». URL: <http://webmed.irkutsk.ru/doc/djvu/djvubook.htm> (дата доступа: 15.10.2010).

wine¹⁷ и в GNU/Linux. Программа имеет множество настроек и работает в два прохода. После первого прохода она автоматически предлагает свои варианты «разрезания» электронных страниц для удаления черных полей и центральной полосы. Далее необходимо вручную просмотреть все отсканированные страницы и внести необходимые коррективы. Краткое практическое руководство по использованию этой программы и другую документацию к ней также можно найти на сайте кафедры библеистики МПДА¹⁸.

Создание единого электронного файла отдельной журнальной статьи с новым титульным листом

Программа ScanKromsator в качестве результата своей работы выдает ряд файлов в формате tiff. В настоящее время в качестве наиболее распространенных форматов для отсканированных книг используются pdf¹⁹ и djvu²⁰. Нами в качестве основного формата используется pdf, хотя конвертировать его в djvu средствами GNU/Linux не представляет труда.

Создание единого электронного файла отсканированной книги в формате pdf из набора tiff-файлов может быть осуществлено различными способами — например, с помощью программ tiffcp и tiff2pdf из пакета libtiff-tools²¹. Все операции производятся из командной строки. Двух команд достаточно для компоновки единого файла, либо написания короткого скрипта²², запускаемого в командной оболочке bash и выполняющего указанные команды²³.

На наш взгляд, правилом хорошего тона является наличие у результирующего pdf-файла подробного библиографического описания — это понадобится, когда файл будет скопирован каким-либо пользователем из электронного архива в отдельную директорию на его компьютере и начнет «самостоятельную» жизнь. Мы считаем уместным, чтобы оно

¹⁷ Официальный сайт: <http://www.winehq.org>

¹⁸ Сергеев В., свящ. Обработка отсканированных книг в программе ScanKromsator 5.6A: краткое практическое руководство // Сайт кафедры библеистики МПДА. URL: http://bible-mds.ru/soft/scankromsator/1-sergeev_v-scankromsator-start-guide.html (дата обращения 15.10.2010).

¹⁹ Формат pdf (сокр. от англ. Portable Document Format — переносимый формат документов) был создан американской фирмой Adobe для того, чтобы документы могли быть одинаково отображаемыми на разных компьютерах и операционных системах; в настоящее время является стандартом для текстовых документов, которые не подлежат редактированию.

²⁰ Формат djvu (читается [дежа-вю]) — открытый формат, разработанный специально для хранения отсканированных изображений.

²¹ Официальный сайт: <http://libtiff.maptools.org>

²² Скрипт — здесь: текстовый файл, при запуске которого командная оболочка bash выполняет описанные в нем команды самой bash либо вызывает указанные в нем программы.

²³ Пример такого скрипта:

```
#!/bin/bash
tiffcp *.tiff tmp.tiff
tiff2pdf -z -x 300-y 300-r o -o article.pdf tmp.tiff
rm -f ./tmp.tiff
```

находилось на дополнительной титульной странице²⁴, что даст определенные преимущества, — можно привести дополнительные сведения, например, о сканировании (указать год, место и название организации, выполнившей его).

Благодаря наличию в GNU/Linux командной оболочки `bash`, этот и следующие этапы могут быть выполнены в автоматическом режиме. Библиограф при этом готовит текстовый файл с подробными библиографическими описаниями статей по номерам журнала, но с небольшими особенностями оформления: в описании выделяются специальным символом имена авторов (например, «+»), номера журналов (в нашем случае — «#») и годы издания (у нас — «@»).

При реализации Санкт-Петербургского проекта был написан специальный скрипт²⁵ для командной оболочки `bash`, который анализирует составленное указанным образом библиографическое описание и сразу выполняет три важные операции:

- генерирует титульные листы (с помощью программы `htmldoc`²⁶);
- сразу соединяет их с файлами в формате `pdf`, при помощи программы `pdftk`²⁷;
- генерирует каталоги в форматах `html` (об этой операции — далее).

Получаемые в результате второго этапа работы скрипта файлы имеют строго оговоренную систему имен: каждое имя содержит год издания, номер журнала и номер статьи в журнале (например, `1869-01-05.pdf`, а в случае двойного номера — `1876-0910-04.pdf`).

Хочу обратить внимание что на этом этапе, благодаря средствам командной оболочки GNU/Linux, все имеющиеся файлы (а их может быть несколько тысяч) обрабатываются в автоматическом режиме, без участия сотрудников отдела сканирования.

Создание электронного каталога

В тех случаях, когда отсканированные книги могут быть объединены по какому-либо критерию (в нашем случае с журналом — либо по номерам, либо по авторам), имеет смысл сразу создавать электронный каталог со ссылками на файлы. Этот каталог должен быть *абсолютно* платформо-независимым, т. е. запускаться в любой операционной системе, в любой программной среде. Было бы несправедливо ориентироваться только на пользователей столь распространенной сейчас в России ОС Windows,

²⁴ В проекте сканирования журнала МПДА «Богословский вестник» библиографическое описание располагается сверху первой страницы статьи; на кафедре библистики МПДА также создавалась отдельная титульная страница.

²⁵ Наш скрипт (загрузить который можно по адресу <http://spbda.ru/script/cr-script.tar.gz>) доступен для использования только под лицензией GPL версии 3.0 или выше.

²⁶ Официальный сайт: <http://www.htmldoc.org>

²⁷ `pdftk` — сокр. от англ. «Pdf Toolkit»; официальный сайт: <http://www.pdfhacks.com/pdftk>

игнорируя при этом не только перспективную систему GNU/Linux, но и систему Mac, весьма популярную сегодня.

Самый простой и самый эффективный способ решения этой задачи — создание каталога в формате html, т. е. в виде файла, который может открыть *любой* браузер интернета. Широкая распространенность этого формата в компьютерном мире позволит открывать файлы данного типа на протяжении многих лет на разных электронных устройствах. Кроме этого, подготовка каталога в указанном формате позволит быстро размещать электронный архив в интернете без дополнительной доработки.

Генерацию каталогов выполняет на третьем этапе своей работы описанный выше скрипт для командной оболочки bash. Обходя файл библиографического описания, он автоматически генерирует два электронных каталога в формате html — по номерам журналов и по авторам, создавая при этом ссылки на файлы с учетом их строго оговоренных наименований.

Некоторые итоги

Преимущество использования свободного ПО при создании электронной версии дореволюционных книг, на мой взгляд, заключается в следующем: низкой стартовой стоимости проекта благодаря отсутствию платы за ПО, возможности эффективной организации сканирования и обработки и, самое важное, — возможности полной автоматизации как сборки электронных книг (с созданием титульных листов библиографических описаний), так и генерации электронных каталогов. Кроме этого, свободный софт позволяет работать с общераспространенными форматами и обеспечивает кросс-платформенность программных продуктов.

Надеюсь, изложенные в докладе сведения будут полезны при организации проектов по сканированию дореволюционной литературы в различных богословских учебных заведениях.