

## **ЭВОЛЮЦИЯ СИСТЕМ ОЦИФРОВКИ ПЕЧАТНЫХ ИЗДАНИЙ НА ПРИМЕРЕ ИХ ИСПОЛЬЗОВАНИЯ В ПРОЕКТЕ ЭБ «НАУЧНОЕ НАСЛЕДИЕ РОССИИ».**

*Кириллов С.А.*

*(Межведомственный суперкомпьютерный центр РАН)*

Одной из основных задач оцифровки печатных изданий является максимально точный перевод их изображения в цифровую форму, с учетом лучших достижений современной техники и уровня развития программного обеспечения на данный период времени.

При выборе оборудования для оцифровки изданий необходимо учитывать как многообразие источников для оцифровки (книги, журналы, газеты, манускрипты и др.), так и специфику их качественных параметров: форму и размер, материал, из которых они изготовлены, технологию их изготовления (методы печати).

На сегодняшний день основным объектом оцифровки для нашего проекта являются книги и научные журналы, выпущенные с 1760 по 1920 годы. Печатный процесс этого периода, в основной своей массе, монохромен. Это касается не только отображения текстовой информации — печать однородных по цвету (черных) букв на светлой (приближенной к белому цвету бумаге), но и изображений. Изображения (схемы, рисунки, даже реплики картин и фотографий) в книгах этого периода также выполнены по монохромной технологии. Иллюзия изменений градаций серого цвета создавалась с помощью (методом) разбивки изображения в зависимости от его тональности на различные по размеру, но однородные по цвету, штриховые и точечные элементы.

Таким образом, основная масса печатной продукции указанного периода (более 90%), выполненная методом высокой, плоской и частично глубокой печати, не воспроизводит переходы цвета.

Эта особенность технологического процесса печати книг позволила инженерам разработать широкоформатные монохромные (черно-белые) сканеры, которые гос-

подставовали в сканировании библиотечных фондов с конца 1990-х годов по 2007 год. Некоторые из этих моделей, например, модернизированные вариации сканера Minolta PS7000, оказались настолько удачными, что предлагаются под другими названиями и до сих пор (2011 год).

### **Первый этап: 2005-2007 гг.**

Планетарный широкоформатный монохромный Сканер Minolta PS7000, который стал базовым в нашем проекте, позволил качественно перевести в электронный вид более 6000 книг и сшитых документов без повреждения переплета и без необходимости разброшюровки. Сканер Minolta PS7000 был выбран как универсальное устройство, на котором можно обрабатывать различные по форме, размеру и толщине документы. Максимальный формат сканера A2 позволяет отсканировать большую газету с разрешением 400 dpi. До формата A3 сканер создает электронные образы с разрешением до 600 dpi. Плавающая колыбель для толстых книг позволяет сканировать оригиналы толщиной до 100 мм. Время сканирования одной страницы формата A4 с разрешением 600 dpi не превышает 8 секунд. Максимальная сменная производительность в нашем проекте — до 900 сканов в день.

Встроенный набор функций по обработке изображения упрощает работу с проблемными книгами. В число таких функций входят: удаление сгиба, маскирование следов пальцев оператора, коррекция параметров изображения, автоматическое определение высоты корешка книги. Эти и другие функции обработки позволяют получать приемлемые изображения в режиме поточного сканирования.

Вследствие отсутствия прижимного контакта, конструкция сканера сводит до минимума механические повреждения оригинала в процессе сканирования, что особенно важно для старых и ветхих изданий.

Опыт эксплуатации сканера Minolta PS7000 показал его высокую надежность в эксплуатации и исключительную эргономичность. Простая панель управления, удобное расположение клавиш ставит этот сканер, по общему

мнению операторов, на первое место по удобству работы, превосходя как все старые, так и многие новые модели планетарных сканеров различных производителей.

*Проблемы монохромного сканирования. Случаи, когда алгоритм бинаризации работает некорректно.*

Тем не менее, эксплуатация сканера Minolta PS7000 выявила и ряд недостатков.

Отсканированный образ страницы преобразуется в бинарный вид, т.е. каждый пиксель представляется одним битом и может быть или черным, или белым. Выполняется это с помощью вычисления порога бинаризации, величины, которая определяет к какому цвету будет отнесен фрагмент изображения (пиксель).

При сканировании качественно отпечатанных на светлой бумаге изданий результаты этого преобразования практически идеальны (более 99,9% символов правильно оцифровываются). Однако, когда величины яркости текста и фона страницы по своим значениям приближаются друг к другу, качество получаемого изображения ухудшается, и при достижении определенного порога получение качественного электронного образа изображения становится невозможным.

*Изменения тона бумаги, не позволяющие получить качественный электронный образ изображения.*

Из нашей практики изменения тона бумажной основы происходят вследствие:

а) Воздействия окружающей среды на бумагу, изменяющие ее фактуру: воздействие света (желтизна), плесень, сильный нагрев или попадание влаги (что характерно для книг, сохранившиеся после пожара) и т.п.

б) Изменения тона участков бумажной основы в процессе использования книги: чернильные, восковые, чайные пятна; различные вариации следов пальцев (засаленные, в чернилах, в саже от свечного или лампового нагара); надписи, закрашивания, подчеркивания или рисунки читателей и многое другое.

в) Особенности технологического процесса изготовления бумаги.

Следует отметить, что практически все книги, выпущенные в период с 50-х годов XIX века по 50-е годы XX века, изготовлены из целлюлозы, содержащей переизбыток кислот [1]. Разрушение бумаги происходит изнутри. Это выражается в изменении цвета бумаги (бумага приобретает коричневый или серый цвет) и в изменении прочностных характеристик бумаги (в худшем случае фрагменты страниц рассыпаются при попытке перевернуть или расправить страницу). Как показала практика, для нашего проекта в особо плохом состоянии находятся книги, выпущенные в период с 1915 по 1929 годы. Вызвано это нарушением технологического процесса в производстве бумаги, а также неудачными экспериментами по изменению самого технологического процесса производства бумаги в период войн, революции и последующей разрухи.

*Влияние типа бумаги на получение качественного электронного образа изображения.*

Отдельный случай, когда трудно получить качественный электронный образ изображения — сканирование изданий, отпечатанных на глянцевой бумаге. Поскольку глянцевая бумага сильно отражает свет, это приводит к тому, что приемный фотоэлемент сканера получает большое число паразитных бликов, что не способствует хорошему качеству сканирования.

*Некачественная печать символов. Некачественное перенесение краски с печатной формы на бумагу.*

Единичные случаи этого типа дефектов встречаются в изданиях всех периодов, однако массовыми они становятся только в период с 1915 по 1929 годы. Объясняется это износом оборудования и нарушением технологического процесса изготовления чернил.

При сканировании на любом сканере, применяющем бинаризацию, этот дефект приводит к тому, что цвет всей отпечатанной на бумаге литеры, либо ее части оказывается слишком бледным и интерпретируется алгоритмом бинаризации в неверную сторону. Человек, в большинстве случаев, свободно читает символы, образованные оттиском типографской литеры на бумаге без краски

— человеческому глазу для этого достаточно уловить малейшие изменения в фактуре бумаги, образованные отриском непрокрашенной литеры. Но почти для всех алгоритмов бинаризации такая задача является неподъемной.

#### *Планшетный сканер Epson GT-15000.*

Таким образом, при оцифровке старых библиотечных фондов часто возникают ситуации, с которыми сканеры, использующие алгоритмы бинаризации, справиться не могут или справляются неудовлетворительно. Для решения этих проблем в комплекс сканирования был включен цветной планшетный сканер Epson GT-15000. Основное назначение сканера: оцифровка карт, иллюстраций, сложных страниц (с которыми не справился Minolta PS7000) с сохранением цветных изображений или изображений в градациях серого в формате TIFF с применением LZW-компрессии.

Другие характеристики сканера позволяют с помощью технологий ColorTrue и Micro Step Drive обеспечивать высокое качество сканирования текста и графики и сканировать документы размером до 297x432 мм (A3). В сканере используется четырехлинейная CCD матрица, обеспечивающая разрешение 600 точек на дюйм. Дополнительно к традиционным линейкам RGB, она содержит четвертую линейку, предназначенную для монохромного (одно- или восьмибитного) сканирования. Эта особенность EPSON GT-15000 обеспечивает повышенную скорость сканирования оригиналов в градациях серого. Сканер оборудован датчиком определения размера оригинала, благодаря которому оператору необязательно выполнять предварительное сканирование, чтобы определить область захвата изображения или вручную выбрать размер документа. Это увеличивает производительность работы и уменьшает вероятность ошибок

#### *Недостатки планшетного сканера Epson GT-15000.*

Недостатки Epson GT-15000 свойственны всем планшетным сканерам: это — небольшая площадь оцифровки (до формата A3) и плохая эргономика, которая приводит к механическому повреждению книги оператором.

## **Второй этап: 2007-2008 г., модернизация сканера Minolta PS7000.**

Одним из вариантов, повышающим как производительность, так и сохранность книг стала модернизация сканера Minolta PS7000 модулем PWB-J, который позволяет сохранять изображения в градациях серого. (Всего для целей проекта приобретено 3 модуля). Однако полностью решить проблему это не позволило, так как цветные изображения все равно приходилось сканировать на Epson GT-15000. Кроме этого все чаще стала обсуждаться тема безопасного освещения при сканировании.

### *Проблема безопасного освещения при сканировании.*

Первоначально (в 1990-начале 2000-х годов) считалось, что в книжных сканерах источник освещения находится на значительном удалении от оригинала и не может повредить его теплом, а минимизация ультрафиолетового излучения просчитывалась инженерами еще при проектировке сканера. Эта задача решалась разработкой или подбором ламп с низкой УФ составляющей в спектре излучения и подбором состава стекла защитного плафона с максимальным поглощением ультрафиолетового излучения. Однако сам принцип бинарного сканирования требует получения максимально контрастного образа. Технологии начала 2000-х годов (как впрочем, и многие современные) достигают этого кратковременным мощным освещением сканируемого оригинала. Так в Minolta PS7000 эту роль выполняет галогенная лампа мощностью 620 ватт, создающая рабочее освещение в 5000 люкс (в других планетарных сканерах рабочее освещение достигает 7000 люкс и более). Примерно с 2002 года появляются работы, которые указывают, что световое облучение не так безопасно, как казалось раньше. Специалисты, производившие лабораторные исследования возможного вредного влияния сканеров на книги, зафиксировали, хотя и незначительную (и не всегда однозначную), зависимость изменения физико-механических свойств всех видов бумаги после сканирования. «Можно констатировать, что сканирование документов не безвредный процесс для бумажной основы документов» — делают осторожный вывод сотрудники НИЦКД РГБ в ста-

тье «Новые исследования оборудования для оцифровывания» [2]. Подводя итог исследованиям, сотрудники РГБ делают следующий вывод, который можно считать постулатом для тех, кто занимается оцифровкой библиотечных фондов: «Оцифровывание следует рассматривать, как стратегию сохранности документов, посредством увеличения доступа к нему в другом формате. Это означает, что цифровая копия должна быть такого качества, чтобы не было необходимости изготавливать повторную копию для нужд микрофильмирования или факсимильного переиздания документов». В качестве практической рекомендации НИЦКД РГБ предлагает ограничить оцифровку каждого библиотечного документа «не более одного–трёх раз».

### **Третий этап: 2008-2010 гг., модернизация комплексов сканирования.**

В связи с вышеперечисленными проблемами для продолжения работ по формированию ЭБ «Научное наследие России» потребовалось новое оборудование, позволяющее:

- а) проводить одновременное сканирование, как черно-белых книг, так и книг, имеющих цветные иллюстрации;
- б) сканировать книги с минимальным освещением;
- в) сканировать ветхие издания, не допускающие разворот на 180 градусов;
- г) сканировать книги большого (до А0) формата.

Для решения первых двух задач были рассмотрены следующие современные модели сканеров: Zeutschel Omniscan, ImageWare Bookeye, сканеры ЭЛАР ПланСкан серии «А» и др.

Характеристики сканеров оценивались по следующим параметрам: возможность потокового качественного сканирования цветных иллюстраций, скорость сканирования, хорошая эргономика, простота обслуживания и управления, стабильность работы и невысокая стоимость обслуживания.

Было решено, что наиболее передовым и интересным техническим решением являются сканеры, предлагаемые фирмой i2S: DigiBook (в России под маркой официально-

го дистрибьютора i2S ЭЛАР ПауэрСкан) и CopiBook (в России под маркой официального дистрибьютора i2S как ЭЛАР ПланСкан-С).

*Комплекс планетарного цветного сканирования ЭЛАР ПланСкан С2-ЦА-600.*

ЭЛАР ПланСкан С2-ЦА-600 это — самостоятельная станция сканирования, в которую уже встроены компьютер, монитор, автоматическая книжная колыбель с прижимным стеклом. Уникальность сканера ЭЛАР ПланСкан в том, что сканирование может осуществляться без дополнительного освещения, что исключает вредное воздействие на ветхие книги. Все необходимые настройки и запуск осуществляются на сенсорной панели. Сканер имеет расширенные возможности настройки и коррекции изображения, установку автофокусирования и автоматические элементы захвата изображения. Благодаря возможности подключения к сети сканер способен сохранять отсканированные образы в любой сетевой папке. Производительность сканера ПланСкан С2-ЦА-600 составляет от 700 страниц в смену, что почти в 5 раз выше планшетного сканера Epson GT-15000 при сканировании цветных изображений.

Технические характеристики. Цветной, формат до А2 (420x594 мм), автоматическая колыбель. Цвет 24(36) бит/точку, разрешение оптическое — 400dpi (на формате А2), 600dpi (на формате А3), толщина оригинала (автофокус) — до 10 см, время сохранения (А2, 400 dpi, цветной режим) — 9 сек, встроенная автоматическая книжная колыбель для оригиналов толщиной до 10 см и весом до 6 кг, прижимное стекло с системой автоматического подъема.

*Комплекс цветного сканирования ЭЛАР ПауэрСкан Д14000.*

Для решения задач крупноформатного сканирования, а также сканирования ветхих изданий, не допускающих разворот на 180 градусов, был выбран сканер модели ЭЛАР ПауэрСкан. Комплекс высококачественного сканирования ЭЛАР ПауэрСкан Д14000 предназначен для бесконтактного сканирования карт, книг, различного рода



скрепленных документов больших форматов, а также прозрачных оригиналов (негативов, позитивов, при наличии модуля подсветки). Комплекс позволяет оцифровывать различные оригиналы с высоким качеством, его модульная конструкция позволяет оцифровывать книги высотой до 30 см и весом до 50 кг. Книжная колыбель, встроенная в раму стола сканера, меняется на плоскую поверхность стола сканирования. Производительность сканера ЭЛАР ПауэрСкан Д14000 составляет 200-600 страниц в смену, также этот сканер позволяет сканировать особо ветхие экземпляры книг (с использованием книжной колыбели) и документы больших форматов.

*Проблемы сканирования при минимальном уровне освещенности.*

Опыт эксплуатации показал, что сканирование с минимальным освещением требует выделения специально оборудованных помещений, в которых можно поддерживать одинаковый уровень освещенности (не менее 300 Люкс) в течение всей рабочей смены.

На качество сканирования при сканировании без подсветки влияют десятки деталей, такие как изменяющаяся яркость освещения в комнате в течение рабочего дня, особенно для помещений, окна которых выходят на южную и юго-восточную сторону, расположение ламп дневного света в помещении, их исправность (мерцание и блики), работа другого сканирующего оборудования. Чувствительные элементы фотоприемника реагируют, например, на коллегу который стоит за спиной сканировщика или проходит рядом со сканером. Поскольку выделить отдельное помещение для каждого сканера не представляется возможным, в 2011 году будет проводиться, в частности, плановая модернизация систем подсветки. В плане эргономики будет выполнено внедрение новых систем крепления, расположения и управления системами подсветки. В плане технологическом — замена люминесцентных источников подсветки светильниками, изготовленными по светодиодной (LED) технологии с уменьшенным излучением в области ультрафиолетового и инфракрасного спектра.

### *Комплекс планетарного цветного сканирования ПланСкан А2-VC-B.*

Одновременно с установкой дополнительных комплексов планетарного цветного сканирования ПланСкан С2-ЦА-600 планируется внедрить новый комплекс планетарного цветного сканирования ПланСкан А2-VC-B. Этот комплекс почти аналогичен по характеристикам ПланСкан С2-ЦА, однако он более компактен, оборудован встроенной светодиодной системой подсветки, легко разбирается и перевозится, поэтому может использоваться как перемещаемая сканирующая станция. Комплекс планетарного цветного сканирования ПланСкан А2-VC-B позволит сканировать оригиналы до формата А2+ с разрешением до 600 dpi, режимы сканирования: черно-белый — 2 бит, серый — 8 бит, цветной — 24 (30) bit/pix; имеет встроенную V — образная книжную колыбель с возможностью раскрытия до 180 градусов, сенсорную русскоязычную панель управления с графическим меню. Может использоваться без управляющего компьютера с просмотром отсканированного изображения на встроенном 18" мониторе.

### **Заключение.**

Таким образом, эволюцию систем оцифровки в проекте ЭБ «Научное наследие России» можно разделить на следующие этапы:

**2005-2007 гг.** Ввод комплекса сканирования на основе монохромного планетарного сканера MINOLTA PS7000 в связке с цветным планшетным сканером Epson GT-15000.

**2007-2008 гг.** Комплексная модернизация сканера MINOLTA PS7000, в частности, модулем PWB-J, позволившая повысить качество и безопасность сканирования, увеличить производительность труда, производить сканирование печатных материалов, как в черно-белом изображении, так и в градациях серого.

**2008-2010 гг.** Переход на цветные широкоформатные книжные сканеры. В качестве базового варианта был принят комплекс цветного сканирования на основе сканера ПланСкан С2-ЦА-600. Для сканирования печатных материалов особо больших размеров до формата А0

был внедрен один комплекс цветного сканирования на основе сканера Пауэрскан Д14000 А0-20/25.

Переход на цветные широкоформатные книжные сканеры позволил:

- улучшить качество сканирования;
- снизить трудоемкость выполнения работ по получению сканов с различных типов печатных изданий, т.е. мы получили производственную схему, по которой одну книгу сканирует один оператор на одном сканере;
- повысить безопасность сканирования за счет исключения механического повреждения оригиналов при сканировании на планшетном сканере и физического повреждения за счет снижения мощности освещения сканируемого оригинала.

**2011 г. и планы на будущее.** Главная задача настоящего периода — это полный переход на цветные широкоформатные книжные сканеры, обеспечивающие максимально точный перевод печатных изданий в цифровую форму. Базовым вариантом остается комплекс цветного сканирования на основе сканеров серии ПланСкан С2-ЦА. Увеличивается количество сканеров, производится модернизация отдельных элементов комплекса, разрабатываются новые алгоритмы работы оператора, изменяется технология сканирования. Для решения задач мобильного сканирования среднеформатных изданий будет введен комплекс цветного сканирования ПланСкан А2-УС-В.

#### *Литература*

1. *Patkus, Beth. Assessing Preservation Needs, A Self-Survey Guide. — Andover : Northeast Document Conservation Center, 2003. — P. 46.*
2. *Новые исследования оборудования для оцифровывания / О.И. Перминова, Е.Д. Яхнин, Т.И. Степанова, И.В. Бурцева, А.Н. Каменский, В.Н. Попунова, А.И. Шарикова // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества. — М., 2002. — Т.1. — С. 243-247.*