

Краевое государственное бюджетное научное учреждение культуры  
«ДАЛЬНЕВОСТОЧНАЯ ГОСУДАРСТВЕННАЯ НАУЧНАЯ БИБЛИОТЕКА»

Хабаровский региональный центр с доступом к ресурсам  
Президентской библиотеки имени Б.Н. Ельцина

**СОВРЕМЕННЫЕ ТЕХНОЛОГИИ В ОБЛАСТИ ОЦИФРОВКИ  
РЕДКИХ И ЦЕННЫХ ИЗДАНИЙ КРАЕВЕДЧЕСКОГО ХАРАКТЕРА**

Методические рекомендации

Составитель И. К. Сорокин, заведующий отделом ХРЦПБ

Хабаровск

2014

## Содержание

Вступление.....	3
1. Общие сведения .....	4
1.1. Оцифровка .....	4
1.2. Сканирование .....	4
1.3. Сканер .....	4
2. Технологии оцифровки .....	4
2.1. Сканирование оригинала.....	5
2.2. Обработка сканированных изображений .....	7
2.3. Преобразование сканированных изображений.....	8
2.4. Типы электронных копий .....	10
2.4.1. Мастер-копия .....	10
2.4.2. Пользовательская копия высокого или среднего разрешения .....	10
2.4.3. Служебная копия низкого разрешения .....	10
Заключение .....	11

## Вступление

Каждый год все меньше людей находят время для посещения библиотек с целью поиска информации, а чаще используют поисковые системы, электронные каталоги, электронные библиотеки. Чтобы стать важным источником знаний библиотека должна быть доступной и оперативной. Также библиотека должна выполнять и другую функцию – сохранить первоисточник от разрушения.

Некоторые из этих ценных и редких документов существуют в единственном экземпляре, однако востребованы читателями, исследователями и учеными. Правда, время и человеческий фактор негативно сказываются на физическом состоянии этих материалов, а поэтому доступ к ним очень ограничен. Использование же оцифрованных книг в читальном зале библиотеки или через сеть Internet помогает сохранить оригиналы, защитить от полной потери, одновременно удовлетворяя потребности пользователей в доступе к ним.

Одним из важнейших направлений развития современной информационной сферы является оцифровка культурного наследия, позволяющая не только сохранить культурные ценности для последующих поколений в новых форматах, но и сделать их более доступными для широкого круга пользователей<sup>1</sup>. Возможность получить доступ к редким и ценным историческим документам является обеспечением одного из фундаментальных конституционных прав человека – права на информацию.

Цель работы – дать общую характеристику современных технологий в области оцифровки редких и ценных изданий

Работа предназначена для сотрудников Дальневосточной государственной научной библиотеки.

---

<sup>1</sup> Грузнова Е.Б. Стандарты оцифровки: зарубежный и отечественный опыт // Президентская библиотека. – 2012. – Сборник научных трудов. – С. 95.

## **1. Общие сведения.**

### 1.1. Оцифровка

Оцифровка - процесс перевода документов в цифровую форму, включает в себя получение копий страниц в виде графических изображений, осуществляемое путём сканирования с последующей обработкой и сохранением в одном из форматов графических файлов, формирование электронных копий документов в постраничном представлении со сквозной нумерацией файлов-имиджей страниц.

### 1.2. Сканирование

Процесс получения любого изображения (страницы книги, слайда, фотографии и т. п. ) в электронном виде, пригодном для дальнейшей обработки с помощью программного обеспечения персонального компьютера. Осуществляется при помощи специального устройства - сканера.

### 1.3 Сканер

Сканер — устройство, выполняющее преобразование расположенного на плоском носителе (чаще всего бумаге) изображения в цифровой формат.

## **2. Технологии оцифровки**

Оцифровка объектов библиотечного хранения преследует следующие основные цели:

- повышение качества обслуживания пользователей: становится возможным доступ пользователей к тем объектам хранения, оригиналы которых находятся в неудовлетворительном состоянии и не выдаются по запросам

- улучшение сохранности оригиналов: уменьшается износ оригиналов, т.к. оцифровка производится один раз и в дальнейшем пользователям предоставляются цифровые копии

- возможность восстановления оригиналов полиграфическим способом их в случае утраты : при условии соблюдения несложных контрольно-профилактических мер цифровая копия обладает неограниченным сроком хранения

- облегчение распространение информации о коллекциях за пределы библиотеки : цифровые копии позволяют расширить возможности межбиблиотечного обмена.

Оцифровка книг предполагает несколько этапов:

- сканирование оригинала книжного издания
- обработка сканированных изображений с помощью программного обеспечения ( графического редактора)
- преобразование сканированных изображений в один или несколько электронных форматов с целью получения одной или нескольких электронных копий книги

## 2.1. Сканирование оригинала

Сканирование книг должно проводиться на специализированных книжных сканерах, которые полностью исключают инфракрасное и ультрафиолетовое освещение оригинала, а также порчу переплета и страниц книг.

Расшивка документов-первоисточников не допускается.

Из всего многообразия моделей книжных сканеров для сканирования документов из фонда редких и ценных изданий оптимальными являются планетарные сканеры.

Планетарный (англ. planetary), он же орбитальный (англ. orbital), он же оверхед- (англ. overhead) сканер — это устройство, где сканирующий механизм расположен над книгой, на некотором от нее расстоянии. Сканеры могут комплектоваться автоматическими колыбелями, компенсирую-

щими толщину книги с сохранением горизонтального положения разворота книги, прижимными стеклами, встроенными экранами для просмотра результатов, собственными элементами освещения или без таковых.

Основным элементом планетарного сканера является сканирующая головка, расположенная на высоте нескольких десятков сантиметров над сканируемым объектом. Сканирующая головка может быть устроена по принципу сканирующей линейки и осуществлять сканирование посредством «просмотра» документа от одного края до другого. Также головки могут оснащаться матрицами, устроенными по принципу матрицы цифрового фотоаппарата (фото-сканеры). Такие устройства осуществляют сканирование за одно раскрытие затвора, что значительно ускоряет процесс.

Многие модели сканеров оснащаются книжной колыбелью, которая предназначена для выравнивания высоты страниц сканируемого документа. В зависимости от модели сканера, колыбель может оснащаться прижимным стеклом для разглаживания поверхности документа и уменьшения искажений при сканировании. Для сканирования книг или сшитых документов с ограниченным углом раскрытия существуют V-образные книжные колыбели (ложементы). Стоит отметить, что некоторые производители оснащают сканеры с V-образной колыбелью двумя сканирующими головками, расположенными симметрично и с таким наклоном к вертикальной оси, что каждая головка сканирует только одну страницу книги.

При работе с документами с ограниченным углом раскрытия сканирование производится с применением специальной подставки.

Главное требование к процессу сканирования редких ценных документов заключается в обеспечении максимальной сохранности первоисточников (оригиналов) и недопущении воздействий, ведущих к ухудшению их физического состояния в процессе выполнения работ.

В результате проведения работ по сканированию состав и физическая сохранность документов-оригиналов не должны быть изменены.

Документы сканируются полностью, включая лицевую и тыльную сторону обложки, титульный лист и оборот титульного листа, концевой титульный лист (колофон), оборотную сторону переплётной крышки, вклейки и оборотные стороны вклеек и т.д. Пустые страницы (вакаты) также подлежат сканированию, независимо от наличия или отсутствия на них номера и другой значащей информации.

Встречающиеся в документах страницы, содержащие картинки, планы, схемы, чертежи и т. д., расположенные целно на 2 страницах (развороте), сканируются постранично отдельно.

Настройки яркости и контрастности производятся для достижения наилучшего результата в части чёткости и цветности изображения и для оптимальной эффективности распознавания типографского текста. Оттенки, глубина и насыщенность цвета образов должны быть максимально приближены к оригиналу и максимально единообразны в пределах одной книги.

## 2.2. Обработка сканированных изображений

Для улучшения внешнего вида графических образов после работ по сканированию при необходимости производится дополнительная обработка с помощью графического редактора.

При проведении дополнительной обработки должны быть выполнены следующие требования:

- Строки текста должны быть выровнены по горизонтали.
- Строки текста должны быть без изгибов.
- Все образы страниц одного формата должны быть одинакового размера по высоте и ширине.
- Образы большеформатных или мелкоформатных страниц, отличных по размеру от основной части страниц источника сканирования, должны сохранять оригинальный размер.

- На скане должно присутствовать изображение страницы с элементами фона по краям (отступ от края 3-10 мм).

- Программная обработка графических образов не должна приводить к существенным изменениям получаемых образов по сравнению с оригиналом.

После обработки файлу –образу страницы присваивается имя, состоящее из краткого названия издания ( английскими буквами с подчёркиванием вместо пробелов) и порядкового номера образа ( трёхзначного) при последовательном сканировании страниц издания.

В результате должны быть получены графические образы в формате TIFF (LZW compressed), оптическое разрешение 300 dpi или более, в постраничном представлении со сквозной нумерацией файлов-имиджей страниц.

### 2.3. Преобразование сканированных изображений

На основе графических образов, полученных в результате сканирования, формируются электронные копии книги в формате PDF. Данный формат «читается» на любых устройствах, используется для печати и публикации в Интернете, позволяет осуществлять полнотекстовый поиск и навигацию по оглавлению и гиперссылкам, а также обеспечивает защищённый режим с ограничениями на печать и копирование электронной книги.

Файлы (образы страниц в формате TIFF) посредством программного обеспечения Adobe Acrobat Pro объединяются в один файл формата PDF с помощью Инструмента «Страницы» ( «Объединить файлы PDF»).

В зависимости от задач PDF-копии формируются в следующих вариантах исполнения:

- PDF-копия без закладок.

Наиболее простой и быстрый вариант создания электронной книги. Формируется на основе графических изображений отсканированных страниц.



- PDF-копия с закладками по оглавлению.

В результате объединения файлов, в итоговом файле PDF установлены закладки, соответствующие названиям объединяемых файлов-образов страниц. При наличии в издании оглавления, закладки соответствующих оглавлению страниц переименовываются согласно пунктам оглавления и лишние закладки удаляются. Если в оригинале нет оглавления, закладки формируются по тексту вручную: по заголовкам глав, разделов и т.п. Таким образом формируется интерактивное оглавление, которое обеспечивает удобную навигацию по материалу, избавляя читателя от необходимости «пролистывать» страницы в поисках нужного раздела.

- PDF-копия с контекстным поиском.

Данный формат позволяет осуществлять контекстный поиск информации, но просматривать документ в том же виде, что бумажный первоисточник, представляет собой двухслойные PDF-файлы, в которых первый «видимый» слой — это графические изображения страниц с точным воспроизведением текста, иллюстраций, карандашных пометок, а вторым слоем является автоматически распознанный текст без верификации. Файл в формате PDF с помощью Инструмента «Распознавание текста» («В этом файле») преобразуется в файл PDF с распознаванием (возможностью поиска в тексте издания). При этом в Настройках делаются следующие установки:

- Язык распознаваемого текста – «Русский»
- PDF на выходе – «Изображение с поиском»
- Понизить разрешение до – «300 dpi».

## 2.4. Типы электронных копий.

При оцифровке создаются копии документа разного назначения. Страховые функции берёт на себя мастер-копия высокого качества с минимальной обработкой изображения, которая позволяет максимально отразить особенности оригинала, в формате, обеспечивающем полноценное со-

хранение данных , а потому имеющая достаточно большой объём ( как правило это формат TIFF без сжатия или со сжатием без потери качества). Обычно на её основе изготавливают пользовательские копии в облегчённых форматах, имеющие разные степени обработки ( форматы JPG, PDF)

Для объектов хранения в рамках плановой оцифровки возможно изготовление как минимум 3 типов цифровых копий, различающихся по объёму информации в зависимости от назначения и особенностей использования.

#### 2.4.1. Мастер-копия.

Эта копия содержит максимально возможное количество информации. Она используется для восстановления объекта в случае его утраты, для других полиграфических целей, для некоторых видов исследований и как основа для изготовления других типов цифровых копий. К файлам, содержащим архивные копии, не применяются алгоритмы сжатия с потерями. Данные копии предполагается размещать на носителях с длительным временем существования. В целях охраны авторских прав доступ к копиям данного типа ограничен, а пользователям они предоставляются только в особых случаях. Используется формат TIFF (Tagged Image File).

#### 2.4.2. Пользовательская копия высокого или среднего разрешения.

Эта копия изготавливается из мастер-копии для ценных и редких объектов. Она предназначена для предоставления пользователям в базах данных и электронных библиотеках. Допускается применение алгоритмов сжатия с потерями, но при этом должна сохраняться читабельность текста и возможность различения деталей графических элементов. Объём файла должен соответствовать пропускной способности внутренней компьютерной сети (изображение должно загружаться не более 3-5 секунд). Разрешение этой копии должно быть достаточно низким, чтобы не позволить полиграфическое копирование объекта. Используется формат PDF (Portable Document Format).

### 4.3. Служебная копия низкого разрешения.

Эта копия изготавливается из мастер-копии или пользовательской копии. Она предназначена для выставления на web-сайте и для предварительного просмотра при поиске по локальной компьютерной сети, в том числе для предоставления пользователям в базах данных и электронных библиотеках. Допускается применение алгоритмов сжатия с потерями, уменьшение разрешения графического образа, внедрение в документ дополнительных данных (например, водяных знаков, метаданных и т.п.). Объем файла должен соответствовать пропускной способности шлюза в Интернет. Используется формат JPEG (Joint Photographic Experts Group).

Стандартным цифровым носителем для размещения и хранения цифрового и библиографического контента являются оптические DVD компакт-диски. Диск должен соответствовать стандарту «Оранжевая книга, часть II» и требованиям ГОСТ 27667-88, ГОСТ 28376-89, ГОСТ Р 51121-97. Сессия записи на каждом диске должна быть финализирована, то есть закрыта.

По окончании записи должна быть выполнена процедура верификация данных и проверка диска на отсутствие компьютерных вирусов.

### **Заключение.**

Оцифровка документов – дело дорогое, трудоёмкое, но необходимое для сохранения библиотечных фондов и обеспечения доступа к информационным ресурсам библиотек.