

Рекомендации по оцифровке материалов из фондов библиотек

Утверждены 23.10.2013 г.

Основные положения

Оцифровка объектов библиотечного хранения преследует следующие основные цели:

- * Повышение качества обслуживания пользователей.
- * Улучшение сохранности оригиналов и восстановление их в случае утраты.
- * Облегчение распространение информации о коллекциях за пределы библиотеки.

Создание цифровых электронных копий объектов хранения предоставляет следующие преимущества:

- 1) Уменьшается износ оригиналов, т.к. оцифровка производится один раз и в дальнейшем пользователям предоставляются цифровые копии. Становится возможным доступ пользователей к тем объектам хранения, оригиналы которых находятся в неудовлетворительном состоянии и не выдаются по запросам.
- 2) При условии соблюдения несложных контрольно-профилактических мер цифровая копия обладает неограниченным сроком хранения.
- 3) При утрате оригинала возможно его воссоздание полиграфическим способом.
- 4) Возможность распознавания образов (в т.ч. и текста) в цифровых копиях позволяет применять развитые процедуры поиска и группировки, что выводит предоставляемые пользователям услуги на качественно новый уровень.
- 5) Обработка запроса на электронную копию занимает значительно меньше времени, чем на оригинал, что позволяет быстрее обслуживать пользователей библиотеки (например, выдача на экран может быть осуществлена в течении нескольких секунд). В то же время, пользуясь электронными копиями пользователь может за одно и то же время получить значительно больше информации, чем при использовании оригиналов.
- 6) Цифровые копии возможно использовать для распространения информации о коллекциях и в рекламно-коммерческой деятельности библиотеки (создание иллюстрированных каталогов на электронно-оптических носителях, представление коллекций в виде уменьшенных копий изображений на Web-сервере, предоставление файлов издательствам для выпуска печатной продукции и т.п.)
- 7) Цифровые копии позволяют расширить возможности межбиблиотечного обмена.
- 8) Пространственная и спектральная чувствительность используемых для оцифровки сенсоров превышает аналогичные характеристики человеческого глаза, что позволяет проводить над цифровыми копиями некоторые исследования, проведение которых над оригиналами невозможно без специального оборудования. Кроме того, цифровые копии могут подвергаться процедурам компьютерной обработки изображений, некоторые из которых могут быть полезны при различных экспертизах.

Подходы к оцифровке

Обязательный: получение копий страниц в виде графических изображений, осуществляемое путём сканирования или фотографирования с последующей обработкой и сохранением в одном из форматов графических файлов. В этом случае полностью сохраняется оригинальная вёрстка книги и исключается искажение содержания. Возможность поиска по тексту отсутствует.

Поисковый (факультативный): оптическое распознавание текста с последующим сохранением распознанного текста как подложки набора графических символов. В этом случае становится возможен полнотекстовый поиск по книге и индексация больших массивов электронных книг. Для полной идентичности содержания требует ручной верификации.

Типы электронных копий

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.

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.

2) Пользовательская копия высокого или среднего разрешения. Эта копия изготавливается из мастер-копии для ценных и редких объектов, или путём прямой оцифровки для тиражных изданий. 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.

4) 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.). Объём файла должен соответствовать пропускной способности шлюза в Интернет.

Состав оборудования

Все виды оцифровки разделяются на два основных типа: контактную и бесконтактную. Применение контактного сканирования разрешено в том случае, если: а) поверхность оцифровки плоская или может быть распрямлена без ущерба для объекта, б) поверхность оцифровки не может быть повреждена при контакте со стеклом, в) свет устройства не наносит вреда объекту, г) для специальных видов объектов, требующих максимального разрешения (более 1000 dpi, например, слайдов). Во всех остальных случаях должно применяться бесконтактное сканирование.

Осветители, применяемые для беконтактного сканирования, должны быть подобраны таким

образом, чтобы обеспечить качественное освещение объекта, но при этом поток ИК и УФ радиации не должен наносить вред объекту.

Выбор между контактным и бесконтактным сканированием, а также между типами осветителей осуществляется совместно с хранителем объекта и специалистом по консервации.

Пост-обработка цифровых копий, размещение на носителях, хранение и контроль за сохранностью носителей.

Пост-обработка цифровых мастер-копий проводится с целью улучшить качество воспроизведения копии на экране, при этом не должно происходить потери информации, содержащейся в цифровой копии. Диапазон возможных способов обработки копий низкого и среднего разрешения более широк, допускается некоторая потеря информации. Полный состав процедур пост-обработки определяется проводящим ее сотрудником в соответствии с целями дальнейшего использования.

Типовыми действиями программной пост-обработки мастер-копий являются:

- * выравнивание освещенности по полю изображения, если таковое требуется;
- * компенсация изменения цветового баланса, произошедшая при оцифровке, если такая компенсация не была произведена на аппаратном уровне;
- * уменьшение размеров изображения за счет неинформативных полей, если таковое требуется;
- * удаление лишних частей изображения, если таковые имеются (например, край соседней страницы, посторонние предметы, попавшие в кадр и пр.), при этом не допускается потеря информации;
- * расширение яркостного диапазона изображения до максимально возможного;
- * изменение значения гамма (при этом следует ориентироваться на значения гамма, характерные для систем РС);
- * сохранение изображения в файле установленного для мастер-копий формата (при этом сжатие с потерями не допускается).

После обработки и записи мастер-копии на ее основе изготавливаются копии среднего и низкого разрешения. Этот процесс в основном подразумевает следующее:

- * уменьшение размеров изображения путем понижения разрешения, достигнутого при оцифровке, и/или удаления краев, которые признаны ненужными для этой копии, при этом следует следить, чтобы текст на экране оставался читабельным, а детали графики были различимыми;
- * выравнивание изображений по краям (по верхнему краю или по двум верхним точкам) с допустимым углом наклона не более 0,1 градуса;
- * приведение к единому размеру изображений, относящихся к одному изданию. Для изданий, которые имеют большой корешок или неровно расположенные листы, размер графического образа определяется по наибольшему из образов. Поэтому в таких изданиях допускается появление увеличенных черных полей для некоторых образов;
- * сшивка изображений для форматов, превышающих формат поля сканера;
- * уменьшение разрешения (величины, определяющей количество точек на единицу площади);
- * уменьшение размеров изображения в пикселях или сантиметрах относительно оригинального размера (например, 50% от оригинального размера);
- * изменение типа и степени сжатия графического файла;
- * внедрение в документ дополнительных данных (например, водяных знаков, метаданных и т.п.).

При записи цифровых копий на носители предпочтительна группировка файлов, относящихся к

одному объекту, на одном носителе или носителях со смежными номерами. После окончания записи обязательно проводится проверка читаемости изображений с носителя и вычисляются контрольные суммы файлов. Каждому записанному носителю присваивается уникальный номер.

Количество экземпляров мастер-копии зависит от сохранности и уникальности оцифрованного объекта. Копии для объектов, относящихся к старым и редким книгам, рукописям, гравюрам и пр., а также для объектов, повторная оцифровка которых затруднена или по каким-то причинам нежелательна, размещаются в двух экземплярах на различных носителях (дублируется все содержимое носителя, поэтому на одном носителе желательно размещать копии объектов с одинаковым количеством экземпляров). Это позволяет практически исключить вероятность потери копии в случае разрушения носителя.

Состояние носителей проверяется не реже, чем раз в 5 лет. При появлениях признаков разрушения носителей информация переносится на другой носитель, идентичность файлов на двух носителях проверяется с помощью контрольных сумм. Состав процедур контроля состояния носителей разрабатывается совместно со специалистами по консервации.

Параметры оцифровки

При невозможности оцифровки объекта целиком допускается оцифровка отдельными частями (с перехлестом). Части хранятся либо по отдельности, либо помещаются на одно изображение так чтобы между ними оставался просвет в несколько пикселей.

Книги 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.

Развороты вшитых/вклеенных в оригинал листов, на которых расположена логически цельная информация (распашная карта/изображение, несколько изображений под одним названием, таблица с переходящими строками и т.д.), должны быть представлены как единое изображение. Развороты вшитых/вклеенных в оригинал листов, информация на которых логически не связана, должны быть представлены как два изображения. Встречающиеся в книгах страницы, содержащие иллюстрации, расположенные цельно на 2 страницах (развороте), сканируются разворотами. Графические образы по книгам формата, большего, чем поле сканера, содержащие информацию, расположенную в разворот, подлежат сшивке после сканирования.

Текстовые страницы книг сканируются с настройками яркости и контрастности, выставленными для оптимальной эффективности распознавания типографского текста.

Настройки яркости и контрастности производятся для достижения наилучшего результата в части чёткости и цветности изображения и для оптимальной эффективности распознавания типографского текста.

Графические образы должны быть четко сфокусированы (резкость текстовых графических образов должна фокусироваться по тексту). Оттенки, глубина и насыщенность цвета образов должны быть максимально приближены к оригиналу и максимально единообразны в пределах одной книги.

Возможны 2 варианта установки параметров поля оцифровки для сшитых изданий: по границе значимой информации (возможна для современных изданий), либо по границе документа, включая каскад страниц, обложку и минимально фон сканирования (рекомендуется для всех редких и ценных изданий для наглядного представления вида границ документа).

Для страниц одного формата графические образы должны иметь одинаковый размер в пикселях по высоте и ширине. Единый размер должен быть выбран по наибольшему изображению страницы с запасом не более 1–3 мм.

Для варианта обрезки изображения по краю документа с каскадом страниц при приведении к единому размеру поля с фоном могут добавляться в необходимом объеме до размеров наибольшего изображения страницы документа.

Образы большеформатных или мелкоформатных страниц, отличных по размеру от основной части страниц источника сканирования, должны сохранять оригинальный размер. Изображение мелкоформатных страниц допускается как на фоне предшествующей или последующей страницы, так и отдельно (на однотонном черном или белом фоне) с последующей обрезкой по границам мелкоформатного документа. Выбор способа зависит от состава оригинального массива и согласовывается индивидуально.

В результате сканирования книг-первоисточников получаются графические образы в формате TIFF (LZW-compressed) в постраничном представлении со сквозной нумерацией файлов-образов страниц.

После сканирования выполняются процедуры автоматического выравнивания полученных изображений по горизонтали и обрезка в формат издания (с учетом полей).

Не допускается сохранение нескольких изображений в один файл, съемка с разным разрешением в пределах одного объекта.

При оцифровке допускается применение аппаратной коррекции гаммы. Цветокоррекция, индивидуальная программная коррекция изображений не производится.

Цветовой режим:

Режим Grayscale допускается, если на оцифровываемом объекте нет никакой значимой цветной информации. Во всех остальных случаях используется режим Color RGB.

Издания, вышедшие до 1861 г., а также газеты дореволюционного периода и первых лет советской власти, снимаются в цвете (RGB), если нет иного требования фондодержателя.

Оператор сканирования вправе самостоятельно принять решение о необходимости цветной съемки издания в случае плохой контрастности (например, коричневый цвет на желтом фоне).

Текстовые страницы документов, страницы, содержащие черно-белые и серые иллюстрации, а также вакафы, включая оборотные стороны любых иллюстраций, в том числе с печатями, штампами, карандашными и т.д. пометами, привнесенными после выпуска издания (не принадлежат изначально оригиналу), сканируются в режиме 8-bit Grayscale (256 градаций серого). Исключение делается для страниц, имеющих авторские или владельческие пометы (посвящения, автографы) или ярко выраженный оттенок цвета, отличный от основного массива страниц оригинала (титульный лист, вакафы, вклейки и т.д.), и для документов, признанных книжными памятниками (полностью цифруются в режиме 24-bit RGB). Обложка издания и отдельные страницы, содержащие цветные иллюстрации или иллюстрации серого цвета, но выполненные на бумаге, отличной по фактуре от основного массива страниц, сканируются в режиме 24-bit RGB (цветной режим).

Разрешение мастер-копии:

- 1) фотоальбомы и фотографии: максимально возможное в диапазоне 300-700 dpi;
- 2) текст с рисунками и без: 300-450 dpi;
- 3) гравюры: 300-1200 dpi;
- 4) карты, атласы: 300-700 dpi.

В случае необходимости отдельные изображения фотографий/гравюр/карт могут быть сохранены в файл.

Форматы файлов

В качестве основных форматов представления цифровых образов отсканированных материалов должны использоваться форматы TIFF, JPG и PDF.

При сканировании использовать TIFF с LZW-компрессией, оптическое разрешение 300 dpi или более.

Для пользовательской и служебной копии использовать JPG с компрессией сохранением 80% качества и, при необходимости, понижением разрешения.

Для представления копии в виде электронной книги должен использоваться формат многостраничного PDF. Файл PDF должен быть оптимизирован для Web и собран из копий в формате jpg с допустимыми для данного типа копии (пользовательская, служебная) параметрами сжатия и разрешения.

Название формата и цель использования	Программы, которые могут открывать файлы	Рекомендуемый метод сжатия
TIFF Tagged Image File Format для мастер-копий	Большинство растровых редакторов и настольных издательских систем; векторные редакторы, поддерживающие растровые объекты	Без сжатия LZW (сжатие без потери качества)
JPEG Joint Photographic Experts Group для пользовательских и служебных копий	Большинство графических программ, Интернет-браузеры	JPEG (можно выбрать степень сжатия. Не рекомендуется сжатие более 50% от оригинала)
PDF - Portable Document Format для пользовательских и служебных копий	Межплатформенный формат файла, переносимый на компьютеры любых систем. Для просмотра файла нужна программа Acrobat Reader или аналог, в т.ч. плагины браузера	LZW, JPEG

Требования к доступности информации

Файлы должны открываться на просмотр стандартными средствами, предназначенными для работы с ними в среде операционной системы, начиная с MS Windows 2000/XP без предварительного вывода на экран каких-либо предупреждений или сообщений об ошибках.

Не допускается в файлах устанавливать опцию запрета печати содержимого файла.

Не допускается устанавливать в файлах парольную защиту на открытие файла. Шрифты, иллюстрации и другие файловые объекты должны быть встроены («внедрены») в тело PDF-файла.

Гипертекстовые ссылки из одного PDF-файла на другие файлы не допускаются. При копировании файлов с носителей информации, просмотре на компьютере или распечатке на устройстве печати не должно возникать ошибок, связанных с некачественной подготовкой или записью на носитель, физическим повреждением или браком в носителе. Подробные сведения о процедурах, применяемых при сканировании аналоговых документов, а также перечень характеристик выходных файлов, используемых для идентификации и регистрации цифровых копий, должны быть включены в Технологическую инструкцию по оцифровке.