

Евтехов Роман Артурович  
сотрудник цифровой лаборатории, научной библиотеки  
Бурятского государственного университета  
аспирант  
E-mail: [lib-lab-scan@mail.ru](mailto:lib-lab-scan@mail.ru)  
Тел.: 89085920169  
Адрес: г. Улан-Удэ, ул. Гагарина 54-7, индекс 67 00 24

### **Формирование электронного контента: новые условия и стандарты качества.**

**Аннотация:** В статье рассмотрен новый этап развития внедренных в библиотечную деятельность вуза новейших инновационных разработок. Серьезное внимание автором уделяется изменениям стандартов качества и требованиям к проведению оцифровки в условиях внедрения новых условий образования и науки.

**Ключевые слова:** электронная библиотека, контент, оцифровка фондов, стандарт, антиплагиат, цитирование.

### **Formation of content Digital libraries: new conditions and quality standards.**

**Annotation:** The article describes the new stage of development embedded in the activities of the university library of the latest innovations. Serious attention is paid to changes in quality standards and requirements for digitization in terms of introduction of new conditions of education and science.

**Keywords:** digital library, content, digitization, standard, antiplagiат, citation.

Современное библиотечное обслуживание в вузах Республики Бурятия и в целом РФ, как показывают материалы конференции сентября 2014 года прошедшей на базе научной библиотеки Бурятского государственного университета (НБ БГУ), за последние годы претерпело значительные изменения [4]. Автоматизация по большей части оптимизировала «классическую» библиотечную работу по созданию бумажного каталога, хранению и поддержанию формуляров, созданию записей и регистраций карточек, точкованию посещений и т. д. В тоже время развитие процессов информатизации позволяет сократить и традиционную книговыдачу изданий, заменяя её предоставлением доступа к удалённым базам информационных ресурсов и созданием собственных путем оцифровки.

Современные вузовские библиотеки все более преобразовываются в информационные центры, основной их целью является не только сопровождение учебного процесса и обеспечение информационных потребностей всех посетителей, но и способствование научной деятельности вуза. Одновременно с этим процессом усовершенствования идет процесс оптимизации библиотечных кадров, их сокращение, не смотря на то, что в библиотеке появляется намного больше работы связанной с обслуживанием оборудования, работой в интернете, оцифровкой фондов, маркетингом, освоением и трансляцией новых информационных ресурсов, возможностей и технологий, оказанием полиграфических услуг и пр.

Эволюцию традиционной вузовской библиотеки можно увидеть на примере НБ БГУ в которой активным образом идет функционирование уже внедрённых инновационных разработок, среди которых: технологии радиочастотной идентификации (RFID) и основанные на них станции самостоятельной книговыдачи, возврата и сортировки [2, с. 17-18]; лаборатория оцифровки фондов с передовым роботизированным оборудованием Kabis II+, программным обеспечением LIMB и Recognition server; ЭБ БГУ на площадке Documentum D2 с автоматизированной загрузкой контента и функцией

поиска по текстам; отдел редких и особо ценных изданий с системой контроля климатических условий хранения и другие.

В связи с тенденцией к развитию вузов РФ как научных центров, сокращению объемов комплектования физическими изданиями библиотек, увеличения нагрузки на студентов в пользу их самостоятельной работы, все это является следствием развития процессов информатизации общества и тенденцией развития западной модели образования, актуальными являются процессы формирования собственного внутриуниверситетского контента. Контент вузов включает по большей части труды преподавателей, сотрудников и учёных вуза (интеллектуальная собственность вуза), а так же материалы фондов особо ценных изданий. Источником пополнения контента ЭБ БГУ являются фонды отделов библиотеки, откуда на время изымаются издания, оцифровка которых не нарушает авторских прав. Вторым источником является издательство БГУ, из которого поступают сформированные перед печатью электронные файлы. Наименьшим источником пополнения контента ЭБ БГУ являются труды, на которые с авторами заключаются индивидуальные договоры о предоставлении права на публикацию в ЭБ.

Все материалы подлежащие оцифровке поступают в цифровую лабораторию и в зависимости от физического состояния, полноты информации и ценности издания проходят оцифровку, графическую обработку и интеграцию в формат pdf., с последующим распознаванием фактически на любом языке (в том числе монгольском, китайском, корейском и т.д.). Весь процесс обработки отсканированного материала производится по методическим рекомендациям и стандартам президентской электронной библиотеки.

Сканер Kabis II+ фирмы KIRTAS позволяет производить сканирование (фотографирование) в автоматическом режиме – с перелистыванием страниц вакуумной «лапой» и бесконтактном режиме – с ручным перелистыванием без применения каких-либо механических частей сканера. Стандартные книги с твердой обложкой сканируются в автоматическом режиме, а книги в мягкой обложке с жестким переплетом и книги фонда редких и особо ценных изданий сканируются только в ручную. Графическая обработка полученных сканов производится в программе LIMB и включает несколько операций по приведению отдельных сканов-изображений к единообразию (в рамках одного издания) и стандарту оцифровки. Затем сформированный документ pdf. переносится в Recognition server, в котором автоматически производится определение языка издания и распознавание [1, с. 22-24]. Итоговый, ужатый для удобства скачивания, снабженный первичными метаданными файл pdf попадает в ЭБ БГУ. Затем специалисты отдела комплектования и учета фондов осуществляют учет и занесение в ЭБ основных сведений об издании, а сотрудники отдела научной обработки фондов обеспечивают издание полным набором сведений и перемещают для общего доступа.

Документы, поступающие из издательства БГУ, как и в большинстве издательств вузов РФ имеют несколько незаконченный вид. Дело в том, что сформированные электронные файлы издательства очень часто не содержат вклеек, которые бывают в итоговом физическом издании и никогда не содержат изображений обложек. Поскольку соответствие физического издания электронному эквиваленту должно быть максимально полным, есть необходимость досканировать недостающие страницы, и формировать итоговый электронный файл аналогичный физическим. Сформированный файл так же распознается и далее передвигается по системе тем же путем что и отсканированные издания. Таким же порядком обрабатываются файлы диссертаций и авторефератов.

Важным отличием ЭБ БГУ является возможность поиска по текстовому запросу, вот почему распознавание является важнейшей операцией производимой при оцифровке. Качество распознавания зависит от трех условий: 1) качество сканирования (модель сканера, разрешение dpi, резкость, цветовой режим и т. д.), 2) качество и разработанность программного обеспечения (форма и степень обработки графических файлов-сканов, словарная поддержка и поколение программ распознавания), 3) исходное качество материала. Вопросы первых двух условий являются исключительно техническими, а вот

на последнем пункте — исходное качество сканируемого материала, хотелось бы остановиться подробнее. Кроме естественных причин, по которым книга может иметь низкий потенциал к распознаванию, таких как выцветание, размытие, повреждение части листа или его отсутствие, поражение грибками и плесенью, существуют, и ряд причин которые можно охарактеризовать как «человеческий фактор». Не секрет что многие особенно популярные издания 50-80-х годов на данный момент являются наиболее повреждёнными изданиями, кроме того читатели любят оставлять разнообразные пометы в книгах, обводить ключевые моменты в тексте, подчеркивать нужные предложения или зачеркивать не нужные. Наиболее «страшными» являются пометки в тексте ручками, маркерами и фломастерами, потому как их удаление без повреждения исходного текста фактически невозможно. Всё это крайне негативно сказывается на конечном итоге распознавания текста. Собственно чтобы уберечь книги от подобных «пометок» и порчи все библиотекари пропагандируют бережное отношение к книгам. Однако, частая работа с электронными ресурсами, полученными путём оцифровки, наводит на мысль, что и сами библиотекари зачастую не прочь «подпортить» книгу в ходе своей работы, в том числе оставить пометки ручкой, но и не только. Например, печать, которая ставится на 17 странице каждой книги, только за редким исключением ставится так, чтобы не задевался текст и номер страницы. Тоже касается печатей на оборотах титульных листов, когда печать часто оказывается прямо в библиографическом описании книги. Схожая ситуация наблюдается при оклеивании издания идентификационными RFID-метками и штрих-кодами. И те и другие наклеиваются на заднюю сторону обложки и нередко частично или полностью перекрывают расположенные там картинки (фото авторов), тексты или описания. С такими явлениями, безусловно, по возможности нужно бороться и самим быть более внимательными и аккуратными, так как подобный «маргинальный» текст или даже картинка после некорректного распознавания (а иного в данном случае быть не может) будет упущен в случае поиска «по запросу» (по тексту).

В этой связи актуализируется методически точная оцифровка изданий с соблюдением номеров страниц, благодаря которым автоматически структурируется книга после интеграции в документ pdf., это делается для осуществления быстрой навигации по файлу. Оцифровка должна осуществляться только в формате pdf, если конечно речь идет о материалах, которые будут доступны как электронные ресурсы. Нужно учесть, что поиск элементарных электронных копий или извлечений из документов уже не актуален. Для обеспечения читателя надежной информацией необходимо чтобы электронная копия документа полностью соответствовала оригинальному физическому изданию. Исследователи понимают насколько это важно, так как для того чтобы дать ссылку на какое либо издание или страницу нужно быть уверенным в том что представленный электронный макет книги является точной копией, текст на страницах не сдвинут, все вклеенные страницы присутствуют, расположение текста соответствует нумерации страниц оригинала и прочее. В противном случае читатель рискует ошибиться при создании ссылок на страницы документа, что уже будет крайне негативно сказываться на его авторитете и может повлиять на конечный результат проверки, на плагиат. Оцифровка библиотечных фондов в исключительно текстовом формате (например, word) на данном этапе просто нецелесообразна. И опять же распознавание текстов формата pdf. должно осуществляться с условием формирования двухслойного pdf. с опцией «текст под изображением страницы», которая позволяет графически видеть исходный материал книги и в тоже время изымать распознанный текст для его дальнейшего цитирования или переработки или поиска.

Пройдя первый этап понимания оцифровки, как проблемы механического копирования, когда основной задачей было сохранение и расширение доступа к изданию, наступил следующий который проходит сейчас не только в нашем регионе, но и целом по РФ. Новый этап - это осознание электронной книги не как набора текстовой информации, но как цельной единицы, которую с максимально доступными возможностями, как

равноценный эквивалент, нужно передать потребителю-читателю. В этом виде электронная книга уже предстает не только текстовой информацией, но становится источником информации сама по себе, что особенно актуально для редких и ценных изданий. И здесь уже поднимаются вопросы форм и границ графической обработки, в том числе уже названные выше – это полное соответствие физического и электронного вариантов и максимально полное распознавание текста. В целом стандарты президентской библиотеки им. Б.Н. Ельцина установленные в широко известных методических рекомендациях оставляют определенную свободу действий для оцифровщиков, но не проясняют некоторые на наш взгляд важные детали. Учитывая новые стандарты образования и нормативы работы электронных библиотек, возникают новые серьезные вопросы, связанные с авторским правом, плагиатом и цитированием.

Одновременно с технологическим совершенствованием НБ БГУ и в дополнение к нему с 2014 года сотрудники библиотеки осуществляют проверку студенческих, магистерских и преподавательских работ на факты наличия плагиата и цитирований в системе «Антиплагиат». Для современного исследователя, который опубликовал результаты своих исследований в физическом варианте, появилась насущная необходимость разместить электронный вариант своей работы в одной из полноценных ЭБ. Почему это важно, дело в том, что системы проверки на неправомерные заимствования являются автоматическими и проверяют механические заимствования (без существенных изменений и последовательностей в тексте). Для того чтобы никто не покушался на такие заимствования необходимо выложить работу на более или менее приличном сайте (сюда конечно не входят любимые студентами: bestreferat, wikipedia и др.). Чтобы любой исследователь, использовавший опубликованную работу неправомерно просто не смог пройти барьер проходного процента оригинальности текста. И в будущем мы возможно сможем избежать историй подобных недавно оглашенной «Диссернетом» со списыванием диссертационных работ депутатами народного хурала РБ (кюн. - Цыбиков Б.Б. – зав кафедрой ВСГТУ, Ботоев Б.Н.). Наилучший вариант для авторов это сайты электронных библиотек, где другие исследователи смогут заметить работу, использовать, процитировать и включить в список использованной литературы, а для этого электронная копия издания должна быть полным аналогом физического издания. На данный момент цитируемость в трудах других учёных это один из важнейших показателей научной значимости работ автора, результативности работы научных организаций и изданий. Важность и актуальность данного аспекта выразилась в появлении целого направления – наукометрии [4, с. 46].

Одним из минимумов учета цитируемости в системе РИНЦ (Elibrary.ru) является авторская научная статья. Разнообразные индексы и подсчёты в системе складываются автоматически из списков правильно оформленной литературы в статьях и монографиях. Здесь опять же необходимо корректировать работу по оцифровке, и вопрос здесь стоит не столько в соответствии общим требованиям оцифровки, сколько в требованиях, предъявляемых к показателям научной работы преподавателей. Во-первых, в связи с учетом отдельных статей как самостоятельных единиц привязанных к сборнику или журналу, в котором они были опубликованы, есть необходимость разделения оцифрованного сборника или журнала на набор статей. В электронную библиотеку должны быть включены оба варианта — сборник (или другое много статейное издание) и набор статей из этого сборника каждый в самостоятельном файле формата pdf. (двуслойный, распознанный аналогично, как и сборник). Что касается магистерских и студенческих сборников статей, то их так же необходимо разделять на статьи, так как не исключено что показатель количества трудов написанных под научным руководством преподавателя в ближайшее время так же начнёт учитываться. Данные критерии опять же стоит применять вкупе с определенными графическими коррективами исходных изображений. Статьи в сборниках и журналах не всегда начинаются с новых страниц, поэтому есть необходимость стирать материалы - начало новой или список литературы

предыдущей статьи, для правильного учёта цитируемости. Так совсем недавно нам приходилось быть свидетелем как один из преподавателей увидев объем своих трудов процитированных другими авторами, решила посмотреть, кто именно и в каком виде это делает. Её очень разочаровал тот факт, что в одной из статей, якобы процитировавших её труд, просто не был удален список литературы, оставшийся от предыдущей статьи. Таким образом, деление многостатейных изданий с удалением лишних элементов других работ учитывая новые требования, предъявляемые к учёту результативности работы ученого, является обязательным требованием к проведению оцифровки.

Вузовскими библиотеками на данный момент накоплен определенный опыт работы с новейшими технологиями в сфере формирования контента электронных библиотек. Необходимость создания собственного, внутриуниверситетского электронного контента с открытым доступом продиктована важнейшей задачей - удовлетворения информационных запросов обучающихся, преподавателей и исследователей, а развитие ЭБ позволяет интегрироваться в общее интеллектуальное пространство региона и РФ в целом. Учитывать изменения условий образования и показателей результативности научной деятельности и преломлять свою деятельность под них есть основное условие развития и дальнейшего формирования из научных вузовских библиотек передовых информационных центров. Эти изменения подталкивают к развитию и соответственно пересмотру стандартов качества, форм и методов оцифровки. В итоге появилась необходимость корректирования некоторых общих и частных моментов при проведении оцифровки фондов и внедрении готовых электронных документов для использования в электронных библиотеках, которые соответствуя новым требованиям и условиям, позволят и дальше развивать сферу науки и образования.

#### Литература:

1. Евтехов Р.А. Деятельность лаборатории оцифровки по пополнению контента электронной библиотеки БГУ / Р.А. Евтехов // Электронные ресурсы и современные технологии вузовских библиотек: решения, инновации, возможности : материалы научно-практической конференции (Улан-Удэ, 8-10 сентября 2014 г.). - Улан-Удэ : Изд-во Бурятского госуниверситета, 2015. - С. 20-26

2. Москвитина Н.А. Автоматизация основных процессов научной библиотеки и внедрение технологий радиочастотной идентификации в НБ БГУ / Н.А. Москвитина // Электронные ресурсы и современные технологии вузовских библиотек: решения, инновации, возможности : мат. науч.-практ. конф. (Улан-Удэ, 8-10 сентября 2014 г.). - Улан-Удэ : Изд-во БГУ, 2015. - С. 20-26

3. Ткачева О.А. Внедрение информационных технологий в научной библиотеке БГУ: из опыта работы / О.А. Ткачева // Современные подходы к библиотечному обслуживанию в образовательных учреждениях : мат-лы науч.-практ. конф. посвящ 75-летию НБ БГУ (г. Улан-Удэ, 7 — 8 сент. 2007 г.). - Улан-Удэ : Изд-во БГУ, 2007. - С. 100-105

4. Электронные ресурсы и современные технологии вузовских библиотек: решения, инновации, возможности : мат. науч.-практ. конф. (Улан-Удэ, 8-10 сентября 2014 г.) / отв. ред. С.В. Васильева. - Улан-Удэ : Изд-во БГУ, 2015. – 80 с.