

Полнотекстовая электронная библиотека – масштабируемое модульное решение

Кедрин Андрей Владимирович, ведущий программист, Институт корпоративных библиотечно-информационных систем Санкт-Петербургского государственного политехнического университета

Соколова Наталия Викторовна, директор, Институт корпоративных библиотечно-информационных систем Санкт-Петербургского государственного политехнического университета

В докладе рассматриваются различные варианты создания собственной электронной библиотеки (внутренней ЭБС). Приведено описание функциональности электронной библиотеки, предложена типизация способов создания электронных библиотек.

В настоящее время значимость создания и владения электронной библиотекой приобретает все большее значение для различных организаций. Так, для общедоступных библиотек в этом плане основным стимулом является необходимость выполнить перевод в цифровую форму части фонда в соответствии с нормативами, определенными министерствами и федеральными целевыми программами. Для организаций сферы образования переход к электронным ресурсам для информационного обеспечения учебного процесса предписан распоряжениями Минобрнауки и Роскомнадзора, поступающими начиная с 2008 года. Тенденция работы с научными ресурсами в электронной среде является одной из основных в мировом научном сообществе, начиная с конца прошлого века [1]. Такими же значимыми являются крупномасштабные национальные и международные проекты в области сохранения культурного наследия в цифровом виде [2].

Таким образом, многие российские библиотеки находятся в настоящее время перед выбором принципиальных решений о реорганизации форм работы с электронными ресурсами. Эти решения необходимо принимать для достижения целей, поставленной министерствами и ведомствами, с одной стороны, и удовлетворения требованиям современных пользователей, с другой.

Электронная библиотека, являясь особым видом информационной системы, включает в себя свойства многих типов информационных систем:

- электронных архивов (долговременной хранение документов в электронной форме, обеспечение функций их быстрого поиска и доставки),
- информационных порталов (Web-интерфейс, современные средства поиска и навигации при работе с документами),
- автоматизированных информационно-библиотечных систем (каталогизация, нормализация данных, учет, средства поиска с привлечением принятых в библиотечном деле лингвистических средств),

- открытых архивов и институциональных репозитариев [3,4,5,6].

Отметим, что на сегодня ни одна из указанных выше платформ не удовлетворяет всему комплексу требований, предъявляемых к современной электронной библиотеке. В этом плане наиболее развернуто требования специфицированы в отношении электронных библиотек для сферы высшего профессионального образования (ВПО), получивших в последние годы название электронных библиотечных систем (ЭБС). Требования можно разделить на контентные, накладывающие ограничения на количественные и качественные показатели размещенных в ЭБС электронных ресурсов, и технологические, предъявляющие требования, в первую очередь, к способам доступа и использования электронных ресурсов пользователями. В данной работе рассматриваются преимущественно технологические требования.

Согласно Приказу Министерства образования и науки Российской Федерации от 31 мая 2011г. № 1975 действующие ФГОС ВПО были дополнены следующим требованием: «Каждый обучающийся должен быть обеспечен индивидуальным неограниченным доступом к электронно-библиотечной системе, содержащей издания учебной, учебно-методической и иной литературы по основным изучаемым дисциплинам и сформированной на основании прямых договоров с правообладателями». Также вводились требования о наличии личного кабинета, о возможности ведения конспекта и другие.

Возможные решения о выборе способа работы с электронными библиотеками определяются особенностями жизненного цикла электронного ресурса. Жизненный цикл включает следующие основные этапы: создание (оцифровка бумажного документа или изначальное создание в электронном виде), описание, размещение в хранилище, использование. Каждый их этапов также включает различные процессы, которые также являются вариативными в способе реализации. К настоящему времени сформировались три основных направления реорганизации работы с электронными ресурсами в библиотеках, а именно:

1. Создание собственной электронной библиотеки (внутренней ЭБС), когда все этапы жизненного цикла библиотека (или организация) выполняет собственными силами;
2. Использование внешних электронных библиотек или ЭБС, когда библиотека вовлечена только в процесс использования электронных ресурсов;
3. Размещение электронных ресурсов библиотеки во внешней ЭБС для дальнейшего использования их на льготных условиях.

В первом случае библиотеки берут на себя полную ответственность и все риски, связанные с контентным наполнением, с развитием программно-технического, организационно-методического и правового обеспечения. Для комплексного решения задач требуется привлекать специалистов в области авторского права, в области обеспечения безопасности в электронной среде и т.д. В целом библиотека начинает выполнять ранее несвойственные ей функции, и это направление требует значительной креативности руководства и хорошей финансовой поддержки на начальном этапе. Однако при этом библиотеки имеют реальный шанс занять устойчивую позицию на рынке информационного обеспечения в долгосрочной перспективе перехода к электронному государству.

Библиотека (или организация) остается не только правообладателем электронных ресурсов, но создателем и владельцем базы данных. Но на библиотеку ложится в том числе и ответственность за выбор технологических решений для электронной библиотеки, что при имеющемся их многообразии требует привлечения грамотных экспертов.

Во втором случае на библиотеки ложится ответственность за адекватный выбор поставщиков электронных ресурсов и заключение договоров. Все затраты связаны с оплатой доступа и/или закупкой электронных ресурсов.

В третьем – библиотека регулирует контентное наполнение, производя сбор (или отбор для оцифровки) электронных ресурсов для передачи их во внешнюю ЭБС, и также контролирует их использование. Библиотеки отказываются от решения собственными силами вопросов обеспечения этапов описания, размещения в хранилище, использования. Эти вопросы переносятся в сферу ответственности внешней организации, которая взамен получает права на использование размещаемых в ее системе электронных ресурсов, в том числе право на предоставление доступа другим потребителям.

Естественно, в ближайшем будущем многообразие вариантов реализации каждого из направлений будет возрастать. Так, в ближайшей перспективе видится размещение электронных ресурсов не у внешнего поставщика ЭБС или агрегатора, а у поставщика «облачного» сервиса. И если уже сейчас библиотеки начинают пользоваться услугами поставщиков «облачного» сервиса типа SaaS (Storage as a Service, сервис хранилища), то скоро будут разработаны специализированные платформы именно для электронных библиотек – PaaS (Platform as a Service) [4]. В этом случае библиотека остается единственным владельцем собственных электронных ресурсов, однако за такой сервис уже необходимо платить внешней организации, провайдеру сервиса.

Помимо этого, библиотека может применять разные из перечисленных направлений за счет распределения требуемых ей электронных ресурсов на разные группы и выбора адекватной формы организации их жизненного цикла. В связи с этим первостепенную значимость приобретают вопросы интероперабельности, совместимости разных систем для организации единой точки доступа ко всему разнообразию электронных ресурсов, которые библиотека предоставляет своим пользователям. Из рассмотренных выше видов систем «прародителей» электронных библиотек принципы обеспечения интероперабельности несколько отличаются. Так, в автоматизированных библиотечно-информационных системах – это протоколы и форматы, в открытых архивах и институциональных репозиториях – свои протоколы и форматы. Причем протоколы и форматы обеих систем уже достаточно устоявшиеся, и имеются распространенные средства отображения их друг в друга для создания гетерогенных распределенных систем информационного обеспечения. Однако эти стандарты охватывают описания (или метаданные), не касаясь формы и формата представления собственно электронного ресурса. В Web-порталах для интеграции в основном используется технология на основе API, часто реализованная в виде Web-сервисов. Такой подход дает принципиальную возможность интеграции ресурса в единую систему, однако по мере увеличения количества электронных

библиотек с собственными (отличающимся друг от друга) API процесс интеграции становится все более сложным. Видимо, данный факт станет стимулом для создания единого API для электронных библиотек хотя бы на ведомственном или отраслевом уровнях.

Ограничения для выбора способа работы с электронными ресурсами для библиотеки часто накладываются доступным объемом финансирования, существующей программно-технической платформой и имеющимся кадровым потенциалом. В настоящее время уже сложились примеры успешной реализации электронных библиотек на основе разных принципов, их совмещения в едином решении. Накоплен опыт развития и трансформации форм работы с электронными ресурсами.

Рассмотрим практический опыт развития электронной библиотеки Санкт-Петербургского государственного политехнического университета (ЭБ СПбГПУ), проследив трансформацию одной формы реализации электронной библиотеки в другую. Отметим, что особенностью ЭБ СПбГПУ является то, что она создавалась в крупном вузе, обладающем значительным научно-образовательным потенциалом и, следовательно, существенным интеллектуальным ресурсом для создания контента электронных библиотек. Кроме того, за счет объединения в едином информационно-библиотечном комплексе нескольких структурных подразделений, поддерживающие разные этапы жизненного цикла электронного ресурса, была обеспечена возможность эффективного решения всех задач, связанных с вводом в эксплуатацию электронной библиотеки. Каждое из подразделений имеет свою сферу ответственности, а именно:

- Фундаментальная библиотека отвечает за отбор произведений для оцифровки и сбор произведений в электронном виде с подписанием соответствующих документов, описание и размещение электронных ресурсов;
- Институт корпоративных библиотечно-информационных систем выполняет разработку и внедрение программного обеспечения, сопровождение серверов для функционирования программного обеспечения и хранения электронных ресурсов;
- Центр цифровой репрографии выполняет оцифровку произведений из фонда библиотеки.

Все подразделения совместно разрабатывают технологии управления электронной библиотекой и соответствующую документацию. Разработка программного обеспечения происходит в тесном контакте с компанией «Открытые Библиотечные Системы» (<http://obs.ruslan.ru>). По сути информационно-библиотечный комплекс является площадкой для внедрения новейших разработок. В области договорной работы и лицензирования электронных ресурсов содействие оказывают юридические службы университета.

Несмотря на указанные особенности электронной библиотеки СПбГПУ, процесс развития является характерным для многих библиотек, во всяком случае, крупных. Выполненные разработки могут быть широко использованы в вузовских и публичных библиотеках, а также в объединенных электронных библиотеках консорциумов и региональных проектов.

Итак, в процессе развития электронной библиотеки СПбГПУ можно выделить три основных этапа реализации электронной библиотеки собственными силами, соответствующие решаемым на каждом этапе задачам, а именно:

- 1 этап. Становление электронной библиотеки, начальное накопление контента.
- 2 этап. Модернизация технологий сбора ресурсов в электронную библиотеку и совершенствование технологической базы хранения и использования электронных ресурсов.
- 3 этап. Создание «продвинутых» форм интерфейса пользователя и дополнительных сервисов. Интерфейс электронной библиотеки – отдельный сайт.

На каждом из этапов возможно устойчивое функционирование системы, решается вполне определенный круг задач. Таким образом, можно говорить о трех вариантах создания собственной электронной библиотеки:

1. Электронная библиотека с минимальным набором специфических функций.
2. Электронная библиотека, удовлетворяющая технологическим требованиям ЭБС. Интерфейс электронной библиотеки – отдельный сайт.
3. Электронная библиотека, удовлетворяющая технологическим требованиям ЭБС. Интерфейс электронной библиотеки интегрирован с сайтом библиотеки.

Единым принципом, принятым при создании электронной библиотеки, был принцип эволюционного развития, заключающийся в следующем:

- новые функции, специфические для жизненного цикла электронных ресурсов, не замещают, а развивают функциональность, уже имеющуюся в автоматизированной библиотечно-информационной системе в части этапов описания, хранения (в части метаданных) и использования (в части организации доступа к электронному каталогу и формирования пользовательского интерфейса);
- новые функции добавляются в виде отдельных модулей, что обеспечивает масштабируемость электронной библиотеки, как информационной системы;
- новые функции создаются на основе профильных стандартов (протоколов и форматов) для обеспечения совместимости и интероперабельности с подобными системами.

Кроме того, указанный подход гарантирует возможность формирования единого фонда библиотеки, включающего традиционные и электронные ресурсы. Правомочность принятого подхода о том, что сетевые ресурсы также являются объектами библиотечного учета, подтверждена Приказом Минкультуры России от 08.10.2012 N 1077 "Об утверждении Порядка учета документов, входящих в состав библиотечного фонда", зарегистрированным в Минюсте России только 4.05.2013 (N 28390). Несмотря на то, что пока действие приказа распространяется только на публичные библиотеки, в этом году ожидается принятие аналогичных решений для библиотек сферы образования и науки.

Дополнительно из приведенных общих принципов следует естественная возможность использования единых лингвистических средств и единого поискового интерфейса для пользователя.

Рассмотрим ниже приведенные три варианта создания собственной электронной библиотеки подробнее. Вариант 1 реализуется на платформе АБИС или на платформе Web-сервера, без использования специального программного обеспечения. Файлы размещаются на Web-сервере, доступ к ним осуществляется по протоколам HTTP или FTP. Описания ресурсов размещаются прямо на сайте в текстовом виде или в электронном каталоге библиотеки. Конечно, вариант со списком ресурсов на странице Web-сервера приемлем только для совсем незначительного количества объектов хранения. По мере их накопления следует разработать специализированные функции поиска, которые традиционно присутствуют в АБИС в виде ОРАС (или АРМ читателя). Для доступа к содержанию электронного ресурса используются стандартные средства программы просмотра ресурса данного формата. Для защиты электронного ресурса от несанкционированного использования возможности также ограничиваются средствами стандартных программ просмотра, которые, как правило, являются недостаточно надежными. Такой вариант использовался в СПбГПУ до запуска нового сайта ЭБ «СПбГПУ». Для поиска электронных ресурсов использовался стандартных АРМ читателя АБИС «Руслан», и поиск выполнялся только по библиографической записи на электронный ресурс (рис. 1). Помимо поиска электронных ресурсов в интерфейсе электронной библиотеки, пользователь имеет возможности найти электронные ресурсы в составе фонда библиотеки и через интерфейс ОРАС, или АРМ читателя.

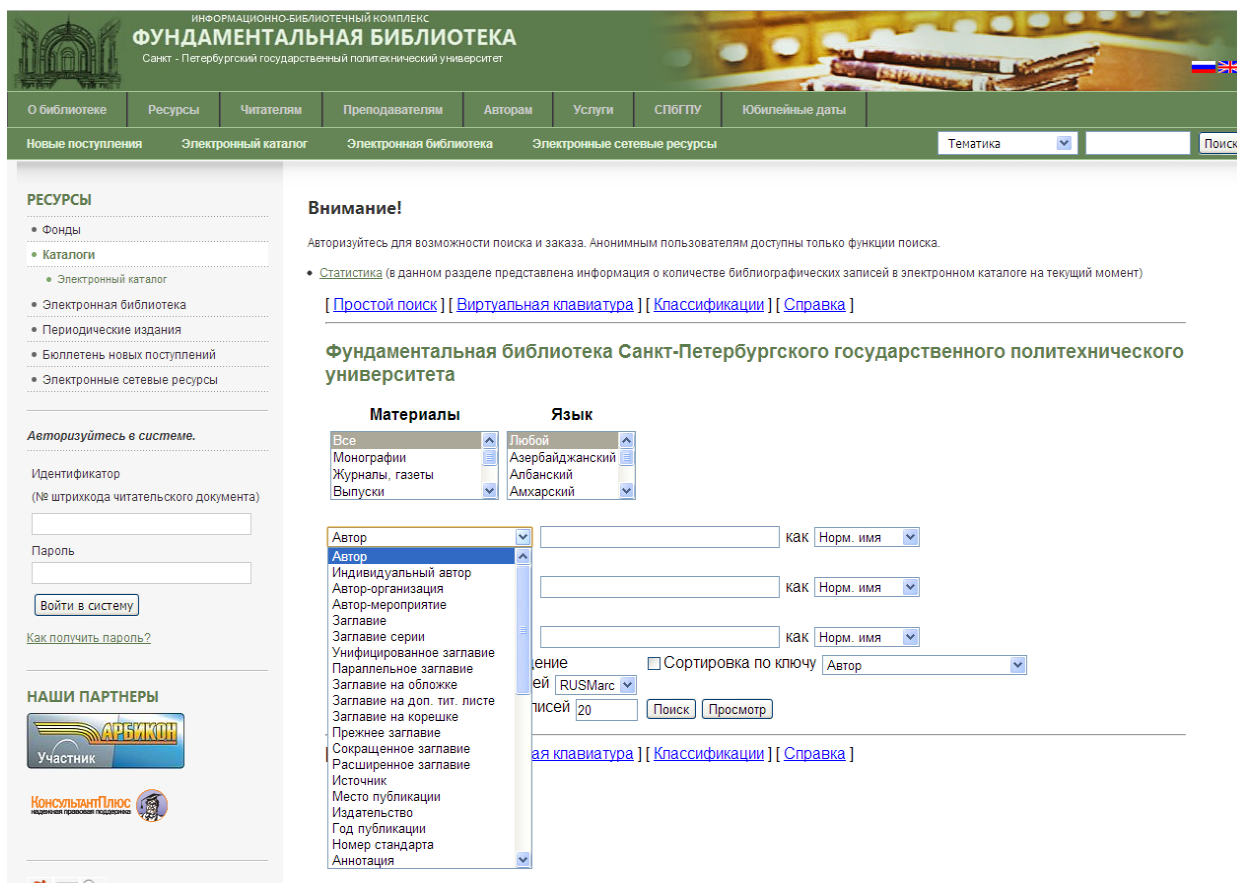


Рис. 1. Страница формирования запроса на поиск в электронном каталоге.

Вариант № 2 предполагает создание отдельного сайта для электронной библиотеки с функциональностью, отличной от традиционного АРМа читателя. В качестве примера можно привести запущенный в эксплуатацию в 2013 году сайт ЭБ СПбГПУ (<http://elib.spbstu.ru/>). Отметим основные особенности данного решения.

Описание электронных ресурсов выполняются с помощью АРМа Комплектования/Каталогизации АБИС «Руслан». Все документы хранятся в файловой системе на сервере. Размещение документов осуществляется сотрудниками библиотеки с помощью стандартных средств операционной системы через сетевые папки. Обработка электронных ресурсов выполняется разными отделами информационно-библиотечного комплекса в соответствии с разработанным регламентом и картой цифрового копирования.

Сайт ЭБ СПбГПУ содержит минимум страниц с текстовой информацией. Главными являются страницы, обеспечивающие поиск и навигацию. Поддерживается поиск по метаданным и полному тексту с учетом морфологии. Разработаны разные варианты поисковых интерфейсов, позволяющие формировать спецификационный поисковый запрос или же применять стратегию итеративного уточнения запроса. На каждой странице сайта есть строка для ввода поискового выражения, есть формы расширенного и атрибутивного поиска (как в АРМе Читателя). Выполняется построение фасетов по заглавию, автору, году публикации, тематике, типу документа и некоторым другим атрибутам. В любой момент времени фасеты отражают срез текущей поисковой выборки по каждо-

му из фасетов, давая объективную информацию для уточнения поискового запроса. Отдельно отметим очень высокую скорость выполнения поиска относительно традиционно используемых в ОПАС-интерфейсах технологий поиска.

Права для доступа к ресурсам определяются для каждого ресурса индивидуально в соответствии с регистрационной формой или лицензионным договором, заполняемыми автором документа, и условия доступа указываются непосредственно в RUSMARC-записи. При этом учитывается местоположение (локальная сеть библиотеки, Интернет и т.п.) и категория пользователя (анонимный, читатель библиотеки, сотрудник библиотеки). На основе этой информации строится матрица условий доступа для каждого объекта хранения электронной библиотеки.

Аутентификация пользователей на сайте ЭБ выполняется с помощью тех же самых логинов и паролей, которые выдаются в библиотеке для электронного заказа литературы и для работы с другими ресурсами на сайте библиотеки.

Для каждого авторизованного пользователя создан личный кабинет, в котором можно составлять списки литературы, смотреть историю своих запросов, формировать закладки для электронных ресурсов.

На сайте ЭБ СПбГПУ также присутствуют описания книг из других электронных библиотек, закупленных у внешних поставщиков электронных ресурсов. В настоящее время это коллекции книг Юрайт и iBooks, причем ресурсы Юрайт загружены непосредственно в хранилище ЭБ СПбГПУ (с указанием продолжительности доступа, соответствующего условиям заключенного договора), а ресурсы iBooks доступны во внешнем хранилище данной ЭБС. Однако система поиска и навигация единая для всех коллекций. Отдельные коллекции могут быть также выделены из внутренних ресурсов ЭБ СПбГПУ.

Интерфейс сайта двуязычный: на русском и английском языках. Сайт создан с использованием технологий ASP.NET MVC4, работает под управлением Web-сервера IIS. Для поиска используется сервер Яндекс.

Сайт зарегистрирован как СМИ и как база данных в Роскомнадзоре. Внешний вид интерфейса электронной библиотекой показан на рис. 2 и 3.

Если электронные ресурсы не доступны для загрузки на компьютер пользователя (нет права на копирование/скачивание и печать), в этом случае для просмотра этих ресурсов используется компонент RBooks. Все передаваемые данные шифруются, что исключает возможность несанкционированного создания копий защищенных ресурсов.

Компонент RBooks является Flash-приложением и поэтому может работать в любом браузере, в любой операционной системе, в которых есть поддержка технологии Adobe Flash 10 (в настоящее время это ОС Windows, Macintosh OS X, Linux и Solaris, браузеры Internet Explorer, Firefox, Safari, Opera, Mozilla, Netscape, SeaMonkey и некоторые другие). Компонент RBooks устанавливается на компьютер пользователя автоматически при открытии в браузере страницы с размещенным на ней электронным документом (при наличии на компьютере Adobe Flash Player 10, который по данным исследований на декабрь 2009 года имеется на компьютерах примерно у 95% пользователей сети Интернет).

Компонент RBooks рассчитан на работу с документами любого размера. Документы небольшого размера загружаются на компьютер пользователя целиком, а при просмотре больших документов на компьютер пользователя предварительно загружаются только оглавление, уменьшенные изображения и тексты всех страниц, а сами страницы загружаются по мере необходимости. Это обеспечивает возможность полноценной работы с документом любого размера через считанные секунды после открытия в браузере соответствующей страницы.

В компоненте RBooks предусмотрены различные варианты отображения страниц документа: последовательно сверху-вниз, одна страница на экране, «книга» с перелистыванием страниц. Также есть поддерживается масштабирование в широком диапазоне, работа с оглавлением, копирование текста в буфер обмена и печать фрагментов документа.

Специальная версия компонента RBooks позволяет контролировать права на работу с документами, т.е. реализует функции DRM-системы. Можно ограничить печать или копирование в буфер обмена. Передача фрагментов электронных ресурсов осуществляется по телекоммуникационным сетям в зашифрованном виде.

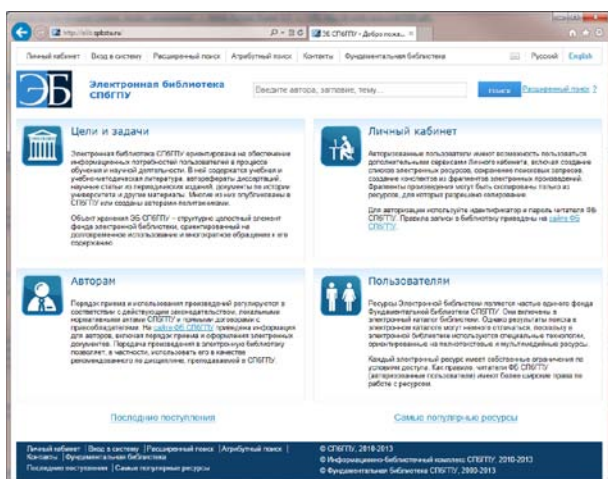


Рис. 2. Главная страница ЭБ СПбГУ

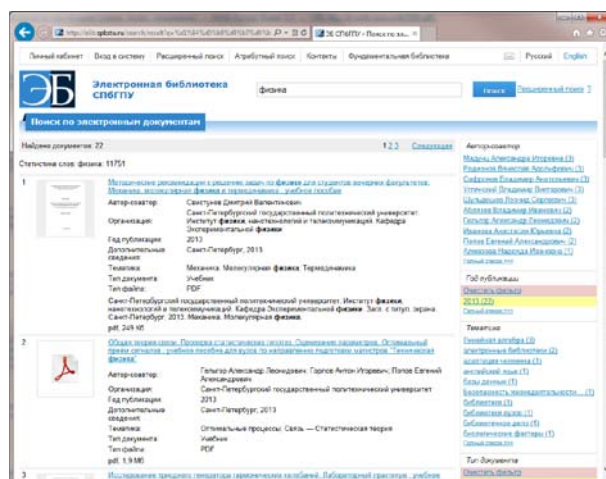


Рис. 3. Страница поиска ЭБ СПбГУ

Отдельно разработана система учета использований каждого объекта хранения ЭБ. Соответствующие статистические данные выводятся на странице детального описания электронного ресурса (рис.4). Также формируются рейтинги популярных ресурсов. Отдельно представлен список новых поступлений.

Вариант № 3 – интеграция с сайтом библиотеки. В качестве примера можно привести раздел «Электронные коллекции» портала Национальной электронной библиотеки Республики Татарстан на сайте <http://kitap.tatar.ru/>. У такого подхода есть следующие особенности:

Описание электронных ресурсов также выполняются с помощью ARMA Комплектования/Каталогизации АБИС «Руслан». Все документы хранятся в файловой системе на сервере. Размещение документов осуществляется сотрудниками библиотеки с помощью FTP-сервера.

Поддерживается поиск по метаданным и полному тексту с учетом морфологии, выполняется построение фасетов. Есть несколько режимов доступа, которые настраиваются индивидуально для каждого ресурса:

- доступ только из локальной сети библиотеки;
- доступ из любой точки Интернета;
- две версии документа, одна из которых доступна для пользователей из локальной сети библиотеки, а другая – для всех остальных пользователей.

Сайт создан с использованием технологий Django, Python, работает под управлением Web-сервера nginx. Для поиска используется сервер Solr.

Внешний вид интерфейса электронной библиотекой показан на рис. 4.

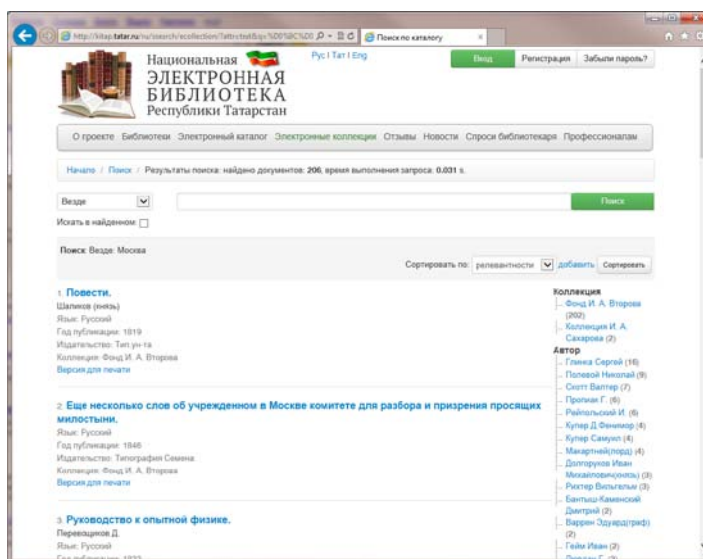


Рис. 4. Страница поиска в Национальной электронной библиотеке Республики Татарстан

Сводные данные по сравнению вышеприведенных подходов к созданию собственных электронных библиотек приведены в таблице 1.

Таблица 1. Сравнение способов создания собственных (внутренних) электронных библиотек

Вариант № 1 – без использования специального программного обеспечения	Вариант № 2 – отдельный сайт для ЭБ	Вариант № 3 – интеграция с сайтом библиотеки
Самый простой в реализации способ, так не требуется покупки, установки и настройки специального программного обеспечения.	Коробочный программный продукт. Быстрое время установки и внедрения. Требуется покупка, установка и настройка спе-	Нетиповое решение. Доработки и внедрение может быть продолжительными. Требуется покупка специального программного

Обычно функциональность такой ЭБ сильно ограничена.	специального программного обеспечения. Богатая функциональность, которая будет доступна сразу после установки и настройки.	обеспечения, а также выполнение работ по интеграции электронной библиотеки с сайтом библиотеки. Функциональность и стоимость определяется тем, что требуется заказчику.
---	---	--

Вариант № 1 обычно используется в организациях на начальном этапе создания электронной библиотеки до того момента, пока не будет принято решение о развитии функциональности и использовании специального программного обеспечения.

Вариант № 2 хорошо подходит в том случае, если ЭБ будет работать как отдельный сайт, а функциональность коробочного продукта полностью подходит или требуются лишь минимальные доработки.

Вариант № 3 позволяет полностью реализовать всю функциональность, необходимую заказчику, но стоимость решения, скорее всего, будет выше, чем при использовании коробочного продукта.

Остается только определиться с требуемой функциональностью и решить, какой вариант лучше всего подойдет для вашей организации.

Список литературы

1. Негуляев Е.А. Открытые электронные архивы и их место в системе информационных коммуникаций [Текст] / Е. А. Негуляев, Е. А. Охезина // Корпоративные библиотечные системы: технологии и инновации: труды IV науч.-практ. конф. АРБИ-КОН, 2 июля-7 июля 2006 г. / Ассоц. регион. библ. консорциумов (АРБИКОН). – СПб., 2006. – С. 46-50.
2. The European Library (TEL) [Электронный ресурс] : [сайт]. – Нидерланды, 2005-2012. – Режим доступа: <http://search.theeuropeanlibrary.org/portal/ru/index.html> (11.06.2013).
3. Соколова Н. В. Электронные библиотеки как факт современного информационного ландшафта [Текст] // Университетская книга. – 2010. – № 2. – С. 62-66 : ил.
4. Дворянчиков В.Я. Библиотеки в облаках - ближайшая перспектива или далекое будущее [Электронный ресурс] / Корпоративные библиотечные системы: технологии и инновации: X Междунар. науч.-практ. конф. и выст. — Электрон. текстовые дан. — СПб., 2012. — Режим доступа: <http://dl.unilib.neva.ru/dl/2576.pdf> (11.06.2013).
5. Соколова Н. В. В поисках единого подхода: методология создания электронной библиотеки для сферы образования и науки [Текст] // Библиотечное дело. – 2009. – № 6. – С. 20-23.
6. Соколова Н.В. Технологии электронных библиотек: шаг за шагом: метод пособие/ Н.В. Соколова, Е.В. Стрелкова, А.И. Племнек; Ассоциация региональных библиотечных консорциумов (АРБИКОН). – СПб.: Изд-во Политехн. ун-та, 2010. – 132 с.