

Программный комплекс ELSA – инновационное предложение для создания электронных библиотек

ELSA Software System – an Innovative Proposition for Building Electronic Libraries

А. В. Кедрин, Н. В. Соколова, Р. Т. Усманов

*Институт корпоративных библиотечно-информационных систем Санкт-Петербургского
государственного политехнического университета, Санкт-Петербург, Россия*

Andrey Kedrin, Nataliya Sokolova, and Rustam Usmanov

*Institute of Cooperative Library Information Systems of St.Petersburg State Polytechnical University,
St.Petersburg, Russia*

В докладе рассматривается масштабируемое решение для создания электронных библиотек. Решение базируется на комплексе свободно распространяемого программного обеспечения. Предусмотрены различные варианты интеграции с АБИС и с другими локальными электронными библиотекам для обеспечения возможности построения распределенных поисковых систем.

К настоящему моменту сложилась устойчивая тенденция перевода традиционных информационных ресурсов в электронную форму, а также создания коллекций электронных ресурсов, существующих только в электронной форме. В эту деятельность уже вовлечено большинство библиотек, и процесс этот развивается практически неконтролируемо. В публичных библиотеках коллекции создаются в основном за счет оцифровки единиц хранения фонда для целей сохранения культурного наследия и/или увеличения доступности часто спрашиваемых ресурсов. В вузовских библиотеках, наряду с оцифровкой редких изданий и учебной литературы, происходит формирование научных и образовательных электронных библиотек, создаваемых силами авторов – сотрудников организации. Однако федеральные и региональные программы создания коллекций электронных ресурсов зачастую не рассматривают библиотеки в качестве полноправных участников таких проектов.

Многолетний опыт развития электронной библиотеки Санкт-Петербургского государственного политехнического университета (СПбГПУ) (<http://www.unilib.neva.ru/rus/lib/resources/elib/>) и консорциума RUSLANet (<http://www.ruslan.ru:8001/rus/consortium/>) показывает, что библиотеки могут и должны совместно создавать и использовать электронные коллекции, сокращая дублирование работ и используя совместно информационное обеспечение нового поколения. Указанные выше электронные библиотеки создавались средствами АБИС, что не позволяло учитывать в полной мере специфику электронного ресурса, имеющего особенности организации его жизненного цикла по сравнению с традиционным печатным документом. Таким образом, актуальной стала задача создания и использования инструмента, предоставляющего более широкие возможности по сопровождению электронного ресурса (включая расширенные способы поиска ресурса по его содержанию) и обеспечивающего при этом возможность объединения, интеграции электронных коллекций различных организаций.

При всем многообразии форм электронных библиотек, существуют типовые этапы их создания и сопровождения. Компания “Открытые библиотечные системы” предлагает открытые решения для создания электронных коллекций, для интеграции электронной библиотеки и АБИС, а также для создания распределенных электронных библиотек.

С технологической точки зрения, электронная библиотека должны обеспечивать поддержание жизненного цикла электронного документа от момента появления проекта документа, его корректировки и согласования с экспертами/ рецензентами/ редакторами и до момента его опубликования в открытом доступе (или другой форме доступа, оговоренной правообладателем) и многократного использования потребителями.

Электронная библиотека должна обеспечить надежное хранение документа, безопасный доступ к нему, а также объединение отдельных электронных ресурсов в различные коллекции, собранные по тематическому или иному принципу.

Электронная библиотека должна быть способна функционировать автономно, решая круг задач, связанных только с обеспечением жизненного цикла электронного документа, а также использоваться совместно с системой автоматизации библиотечных процессов, объединяя и взаимно дополняя их функциональность.

Программный комплекс ELSA (Electronic Library Services and Applications) компании “Открытые библиотечные системы” удовлетворяет всем перечисленным выше требованиям. В качестве основы электронной библиотеки используется свободно распространяемое программное обеспечение и библиотеки программ с открытым исходным кодом, которые позволяют обеспечить следующую функциональность:

1. Хранение электронных документов. Документы могут быть размещены во внутренней базе данных, в файловой системе на сервере, во внешней реляционной базе данных, на удаленных Web и FTP-серверах. Таким образом, каждая организация сможет выбрать наиболее удобный для себя способ хранения документов.

2. Поиск по метаданным и полному тексту документа. Набор метаданных документов соответствует элементам формата Dublin Core: автор, заглавие, аннотация, тематика, ключевые слова и прочие атрибуты. Для организации поиска по полному тексту документа могут быть использованы различные модули лингвистической обработки. Так, например, возможен поиск без выделения словоформ, поиск с учетом морфологии, поиск с обеспечением семантического анализа. В настоящее время в ELSA используется модуль, учитывающий морфологию русского, английского и ряда других языков.

3. Поддержка различных форматов документов. Уже в текущей версии компонента электронной библиотеки поддерживается вариативность форматов представления электронных ресурсов – djvu, pdf, doc, html, ppt, rft, txt, xls. По документу в любом из этих форматов создается полнотекстовый поисковый индекс. В дальнейшем количество поддерживаемых форматов будет постоянно расти.

4. Доступность электронного документа в зависимости от его статуса. Обеспечивается поддержка набора состояний у документа, связанных с его жизненным циклом. В каждом конкретном состоянии определен круг пользователей, которые могут выполнять просмотр документа, его редактирование и изменения статуса.

Общая архитектура системы показана на рис. 1. Особенно следует отметить возможность создания индексов в среде компонента для электронных документов, размещенных удаленно (на любых Web и FTP-серверах, доступных через Интернет). Это позволяет обеспечить быстрый поиск документов во время выполнения запроса на поиск по полному тексту документа.

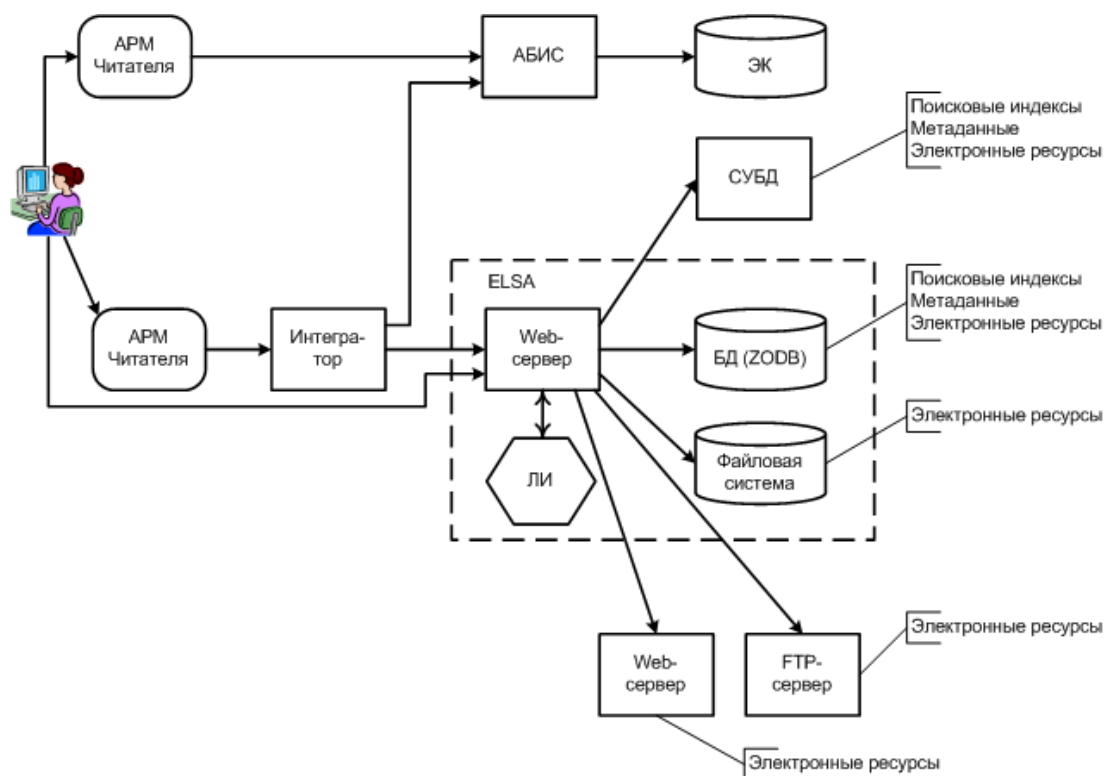


Рис 1. Общая архитектура системы

В электронной библиотеке СПбГПУ содержится около 8 тысяч документов. Библиографические описания всех этих документов создаются и хранятся в АБИС “Руслан” в специальных базах данных. Ранее поиск документов в электронной библиотеке осуществлялся только по метаданным.

В настоящее время в СПбГПУ установлена электронная библиотека ELSA (<http://elsa.unilib.neva.ru>). После установки ELSA появилась возможность осуществлять поиск по полным текстам документам, что позволило пользователям выполнять более качественный поиск необходимой информации.

Внешний вид интерфейса электронной библиотекой показан на рис 2. На левой стороне экрана представлена иерархия электронных ресурсов, которые могут быть объединены в несколько электронных коллекций. В данном случае коллекции связаны с подразделением, выполнившим создание коллекции. Может быть использован и другой принцип объединения ресурсов. При этом один и тот же ресурс может принадлежать нескольким коллекциям.

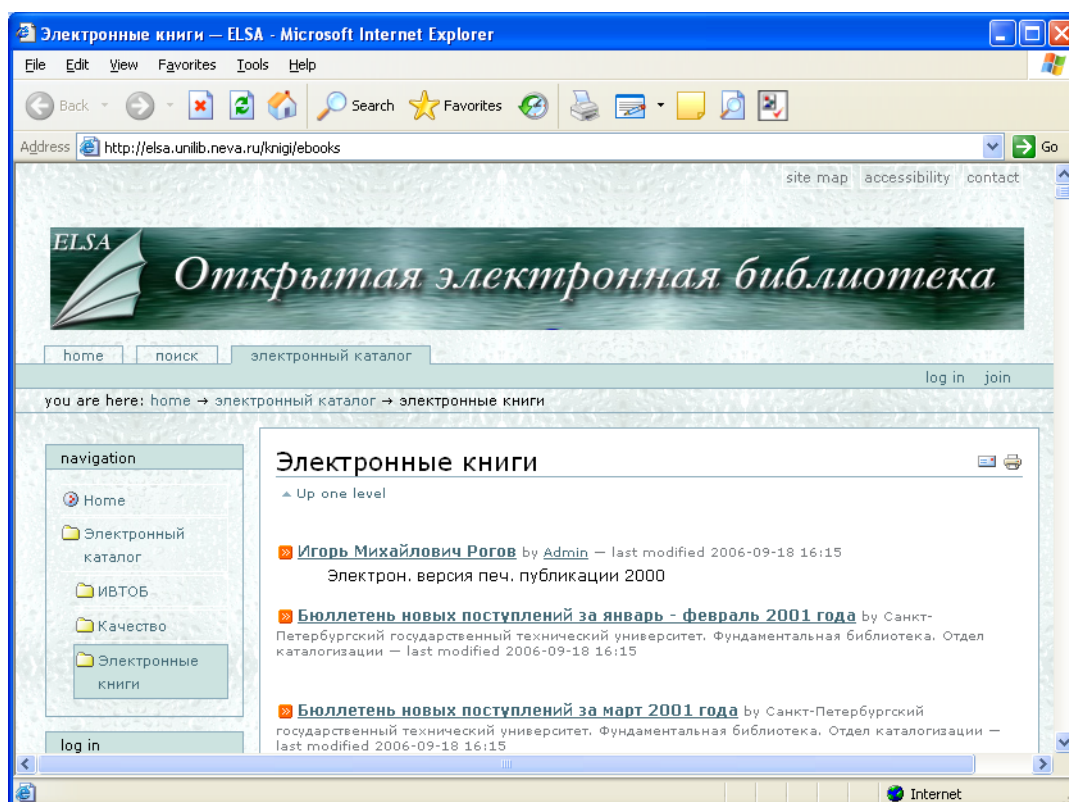


Рис.2. Список документов в каталоге электронной библиотеки

Особо следует отметить возможность интеграции электронной библиотеки ELSA и АБИС. Так, например, разработана специальная программа ELSA-IS для выполнения импорта библиографических описаний документов из АБИС в ELSA и их последующей синхронизации: добавлении новых и обновлении существующих. Эта программа осуществляет выгрузку записей из АБИС по протоколу Z39.50 в формате RUSMARC, их последующую конвертацию в формат Dublin Core и запись в электронную библиотеку ELSA. Применение этой программы в СПбГПУ позволило в течение одного дня выполнить загрузку описаний всех электронных документов в ELSA и дать возможность выполнять по ним поиск. При этом сами документы по-прежнему хранятся на FTP-сервере библиотеки СПбГПУ.

Для обеспечения возможности взаимодействия АБИС и электронной библиотеки разработан дополнительный промежуточный компонент – Интегратор (см. рис. 1). Он обеспечивает при поиске виртуальное объединение массива электронных документов, хранящихся в электронной библиотеки, и электронного каталога, поддерживаемого АБИС.

Таким образом, в обычной форме поиска в АРМе Читателя появляется возможность добавлять к традиционным атрибутам поиска в электронном каталоге еще один – поиск по полному тексту документа. При этом пользователи могут строить поисковые запросы в уже привычном для себя интерфейсе АБИС.

Это решение является целесообразным для объединения электронных библиотек, создаваемых на кафедрах и факультетах или в различных отделах библиотеки, в рамках отдельных проектов, с основным информационным хранилищем в среде АБИС. Все распределенные ресурсы (созданные в различных средах и представленные в различных форматах) становятся при этом частью единого целого, единой системы информационного обеспечения вуза или другой организации. По всему этому информационному массиву можно выполнять поиск через общий интерфейс.

Предусмотрен вариант интеграции и на более высоком уровне, когда происходит объединение различных библиотек. В текущей версии требованием для объединения является представление каталога АБИС по протоколу Z39.50 (т.е. необходимо выставление электронного каталога на локальном или корпоративном сервере Z39.50), а каталога электронной библиотеки – по протоколу SRU. При такой архитектуре распределенной системы мы через единый интерфейс получаем возможность производить параллельный поиск по записям электронных каталогов различных АБИС и электронных библиотек.

В настоящее время подготовлена первая версия дистрибутива ELSA, которую можно загрузить с сайта ОБС (<http://obs.ruslan.ru>).

Компания “Открытые библиотечные системы” продолжает исследования и разработки, связанные как с увеличением функциональности компонента электронных библиотек (качество поиска, удобство управления электронными ресурсами, пр.), так и с совершенствованием процесса интеграции различных компонентов в единую распределенную информационную систему. Работы выполняются в тесном сотрудничестве с СПбГУ и другими партнерскими организациями.