

УДК 681.324

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ СОЗДАНИЯ ЛИЧНЫХ АРХИВОВ ВЫДАЮЩИХСЯ УЧЁНЫХ В ПОРТАЛЕ РАН

А. В. Босов, Р. Б. Чавтараев

Описано решение, обеспечившее пополнение контента Информационного web-портала РАН (www.ras.ru) интереснейшими информационными ресурсами — архивами ряда выдающихся российских учёных. Рассмотрены как нормативно-правовые вопросы создания личных электронных архивов, так и программная технология, разработанная для реализации таких систем и подключения их к информационному portalу.

1. Введение

Российская научная общественность обладает богатейшим наследием, к которому, несомненно, относятся и личные архивы выдающихся российских учёных. В настоящий момент эти ресурсы мало известны и практически не востребованы в силу ряда объективных и субъективных причин. Одна из этих причин — отсутствие информации об архивах и их представления в сети Интернет, уже значительное время являющейся важным источником информации для исследователей.

Личные архивы по сути представляют собой совокупности документов, связанных или имеющих отношение к тому или иному выдающемуся учёному, и тематически объединены в фонды. Эти материалы имеют несомненную историческую и научную ценность, поэтому содержание архива ведётся в соответствии с правилами ведения архивов документов, принятыми в РФ и соответствующим образом регламентированными.

В последние годы в связи с развитием информационных технологий в архивном деле активно применяются различные средства автоматизации, что вполне естественно. В настоящей работе рассматриваются достаточно специфические аспекты архивного дела. Прежде всего, эта специфика связана с характером материалов — личными архивами выдающихся учёных

России. Другая особенность связана с тем, что используется особый вид архивных документов, называемый *страховым фондом*, в котором хранятся копии оригинальных документов, созданные специально для того, чтобы сохранить научное наследие. Использование именно страхового фонда связано, прежде всего, с особенностями нормативной базы архивного дела, принятой в РФ. В Российской академии наук деятельностью по созданию и хранению архивной информации, в том числе и личными архивами учёных, занимается Архив РАН и, в частности, страховыми копиями — отдел страхового фонда (<http://www.ras.ru/MArchive/info.aspx?action=contacts>).

Цель обсуждаемой работы — рассмотреть технологию, обеспечивающую не только сохранность страхового фонда, но и позволяющую представить его на вполне легальной основе в открытом доступе, чтобы дать возможность получить и необходимую описательную информацию, и сами документы всем заинтересованным лицам.

Подобные задачи уже ставились применительно к отдельным частным архивам (см., например, «Архив академика А. П. Ершова», <http://www.erшов.ras.ru>). Однако наша задача — предложить универсальный способ, который, во-первых, должен соответствовать в полном объёме требованиям существующего законодательства в области архивного дела, а, во-вторых, привести к созданию системы, готовой к интеграции в Интернет.

Описанная в данной работе технология позволила создать ресурс — электронный архив «Научное наследие», являющийся, с одной стороны, вполне самостоятельной web-системой, с другой — внешним информационным источником, готовым к интеграции в Информационный web-портал РАН (www.ras.ru), разработка которого ведётся в Академии наук с 2002 г. [1–6].

2. Основная терминология

Основные термины и определения, регламентирующие архивное дело в РФ, приведены в [7, 8]. Из них выделим следующие понятия:

— *архивный фонд РФ* (совокупность документов, отражающих материальную и духовную жизнь её народов, имеющих историческое, политическое или культурное наследие);
— *архивный документ* (документ, сохраняемый или подлежащий сохранению в силу его значимости для общества);

— *архивный фонд* (совокупность архивных документов, исторически или логически связанных между собой).

Являющиеся предметом нашего рассмотрения личные архивы выдающихся российских учёных, таким образом, являются частью архивного фонда РФ, каждый личный архив является архивным фондом, а составляющие его материалы — архивными документами.

Интересно отметить, что нормативы архивного дела не просто допускают применение методов его автоматизации, но и регламентируют их. Так, в [9] указано, что Федеральная архивная служба предоставляет программное обеспечение для автоматизированного государственного учёта документов. Перечислен даже ряд программных комплексов, используемых для учёта («Архивный фонд», «Фондовый каталог»). Ряд архивов переданы в Росархив и конвертированы в базу данных «Архивный фонд» (есть сведения о 132 тыс. фондов по данным на 2000 г.).

Личные архивы учёных, следуя нормативной лексике [10], будем называть *архивными фондами личного происхождения*, так как они состоят из документов, образовавшихся в процессе жизни и деятельности отдельных лиц (*фондообразователей*). Для таких архивов и ставится задача представления составляющих их документов в цифровом виде в Интернет.

Правомочность постановки такой задачи, помимо простой информационной ценности, следующей из исторической и научной значимости документов, объясняется и формальными нормативными положениями [10]. Именно к задачам архивного дела отнесены организации выставок документов, подготовка документов для использования средствами массовой информации, предоставление документов исследователям, т.е. поддержка публичного доступа к архивам. Те механизмы, которые используются для этих целей сейчас, призвать удовлетворительными и эффективными нельзя. Возможность работы с архивами через Интернет позволяет достичь как перечисленных целей, так и многие другие, например информирования научной общественности о наличии и содержании материалов фондов.

Другая решаемая задача — сохранение научного наследия. Именно с этой целью создаются *страховые копии* [11] особо ценных и уникальных документов путём их микрофильмирования. Совокупность страховых копий документов, хранящихся в архиве, составляет *страховой фонд*. Личные архивы выдающихся российских учёных, несомненно, относятся к данной

категории фондов, подлежащих созданию страховых копий, однако непосредственно использовать страховой фонд в каких-либо исследованиях неудобно. Причина этого состоит в том, что под страховой копией в [10] определён негативный микрофильм, а цифровые копии документов в действующих правилах архивного дела не рассматриваются.

Одновременно с созданием страхового фонда создаётся *фонд пользования*. В [11] также указывается, что фонд пользования может создаваться целевым порядком путём создания копий документов на электронных носителях, обеспечивающих идентичность копии и подлинника документа. Таким образом, именно целесообразность создания электронного фонда пользования обосновывает существование архивной системы «Научное наследие».

Другим правовым основанием использования цифровых копий (электронных архивов) является понятие «*научно-справочный аппарат*» [11]. Это структурированная совокупность описаний документов (вторичной документной информации), представленных в различных видах архивных справочников, баз данных, предназначенных для поиска документов и документной информации [11]. Этот комплекс взаимосвязанных и взаимодополняемых архивных справочников, баз данных о составе и содержании документов создаётся на единой научно-методической основе и предназначен для поиска документов и документной информации в целях её дальнейшего использования. К элементам научно-справочного аппарата относятся описи, каталоги, базы данных, указатели, обзоры, исторические справки и т. д.

Таким образом, электронные архивы выдающихся учёных, с позиции нормативно-правовой базы архивного дела, являются элементами научно-справочного аппарата фонда пользования. При этом подключение к электронному архиву средств поиска позволяет решить большинство задач научно-справочного аппарата архива гораздо эффективнее и удобнее для пользователя.

3. Схема электронного архива

Для программной реализации электронного архива был выполнен анализ правил ведения архивов документов с целью формирования необходимых функциональных схем и структур данных. К основным разделам схемы систематизации документов

следует огнесте, прежде всего, понятие архивного фонда личного происхождения [10] (далее, фонда), в который входят документы творческой, служебной и общественной деятельности, личная переписка, документы к биографии, документы о фондообразователе, изобразительные документы, документы, собранные фондообразователем и т. д.

Внутри фондов документы по некоторым признакам группируются в *дела*, которые, в свою очередь, объединяются в *описи*. Систематизированные таким образом в иерархию фонд–опись–дело документы оцифровываются, а затем проводится дальнейшая структуризация полученных информационных массивов. В частности, практика Архива РАН предполагает классификацию дел с помощью словарей «Способ воспроизведения» и «Вид материала», которые включаются в систему с целью поддержки сложившейся в РАН практики ведения страховых фондов.

Основные решения, использованные для построения схемы данных личного электронного архива, приведены в [11]. Понятие *номенклатуры дел* — систематизированного перечня наименований дел — служит основой для составления их описей. Именно в номенклатуре дел предусматриваются: индекс дела (по правилам организации), заголовки дела, количество дел, сроки хранения, примечание. В описательной статье описи приводятся порядковый номер, инвентарный номер, язык, указание на подлинность, копияность, способ воспроизведения, степень сохранности, внешние особенности. Естественно, что для автоматизированной системы ведение ряда атрибутов целесообразно.

Предусматривается дополнительная классификация дел: — по виду материала (словарь «Вид материала»: автореферат, аннотация, гранки, выписка, альбом и т. д.);

— по способу воспроизведения (словарь «Способ воспроизведения»: рукописный текст, автограф, авторизованная машинпись и т. д.);

— язык материала (русский, английский, французский и т. д.).

Следующий шаг в определении необходимых атрибутов даёт понятие *архивной описи* — справочника, содержащего систематизированный перечень единиц хранения архивного фонда/коллекции и предназначенного для их учёта и раскрытия содержания. В неё должны входить: индекс дела, заголовки, крайние даты, количество листов, срок хранения, примечание. В итоговую запись архивной описи включаются также: название

фонда, номер фонда, хронологические границы архивного фонда личного происхождения (даты рождения и смерти физического лица).

В соответствии с принятыми нормативами, электронный архив «Научное наследие» представляет собой набор фондов. При этом каждый фонд связан с конкретным учёным. Фонду соответствует лист фонда, содержащий сведения о местонахождении фонда, дате первого поступления фонда, крайние даты, название фонда, описи дел и др. К описательной части фонда относятся также историческая справка, схема систематизации фонда, методические пособия по обработке фонда и пр.

Ряд атрибутов элементов архива (фондов, описей) в электронном архиве, очевидно, должны формироваться автоматически: реестр описей, крайние даты описей, данные учёта электронных документов, данные учёта использования.

Проведённый выше анализ позволяет сформировать схему данных электронного архива «Научное наследие». Данные архива имеют древовидную структуру, в которой определены сущности: «фонд», «опись», «дело» и «документ». К атрибутам фонда отнесены номер фонда, историческая справка, сведения о местонахождении фонда, сведения о фондодержателе и фондообразователе.

В состав фонда входит произвольное число описей, имеющих номер и номенклатуру дел. Так же, как фонд является корневой структурой для описей, опись является корневой структурой для дел. Каждое дело содержит множество единиц хранения — документов, оптических копий страниц архива.

Дело содержит следующую основную набор архивных атрибутов:

- номер;
- название;
- вид материала (из словаря);
- способ воспроизведения (из словаря);
- крайние даты;
- язык документов (из словаря);
- электронные копии страниц документов.

Все текстовые атрибуты имеют два языковых представления: русское и английское.

Для создания системы хранения и эффективного решения дальнейших вопросов интеграции схема данных электронного архива представлена в форме XML-Schema.

4. Web-система Электронный архив «Научное наследие»

Основной функцией данного архива является предоставление пользователю материалов архива в удобном для него виде за короткой промежуток времени. Для этого разработан Интернет-сайт, поддерживающий следующие функции:

- удобную навигацию по описям с просмотром содержимого описи (списка дел); список дел может быть отсортирован по номеру, названию или виду материала;
- полнотекстовый и атрибутивный поиск (результаты поиска могут быть также отсортированы); в поиске участвуют поля: «номер», «название», «вид материала», «способ воспроизведения», «крайние даты»;
- заказ материалов с доставкой по электронной почте.

Поскольку данные архива являются статичными, функции обновления данных могут быть выполнены стандартными административными средствами платформы хранения данных архива, которой является СУБД MS SQL Server, поддерживающая базу данных на основе представленной выше XSD-схемы. Для хранения электронных копий страниц документов используется файловая система.

Электронный архив представляет собой вполне самостоятельную web-систему, включающую web-сайт и подсистему хранения. Хотя сайт и выполнен в стиле Информационного веб-портала РАН (рис. 1, рис. 2), никаких ограничений на его самостоятельное использование не накладывается.

В качестве платформы приложений использована технология ASP.NET. Кроме того, разработан набор ASPX-шаблонов, обеспечивающих:

- просмотр и навигацию по описям и делам фондов;
- просмотр документов фонда;
- полнотекстовый и атрибутивный поиск по атрибутам дел;
- заказ материалов с доставкой по электронной почте.

В настоящий момент в состав электронного архива «Научное наследие» включены три фонда: почётного академика Н. А. Морозова, академика В. И. Вернадского и академика С. И. Вавилова, ведётся разработка архива академика В. Л. Комарова. Всего в фондах размещено 860 дел и более 210000 документов. Работа по пополнению архивов материалами продолжается и в настоящий момент.

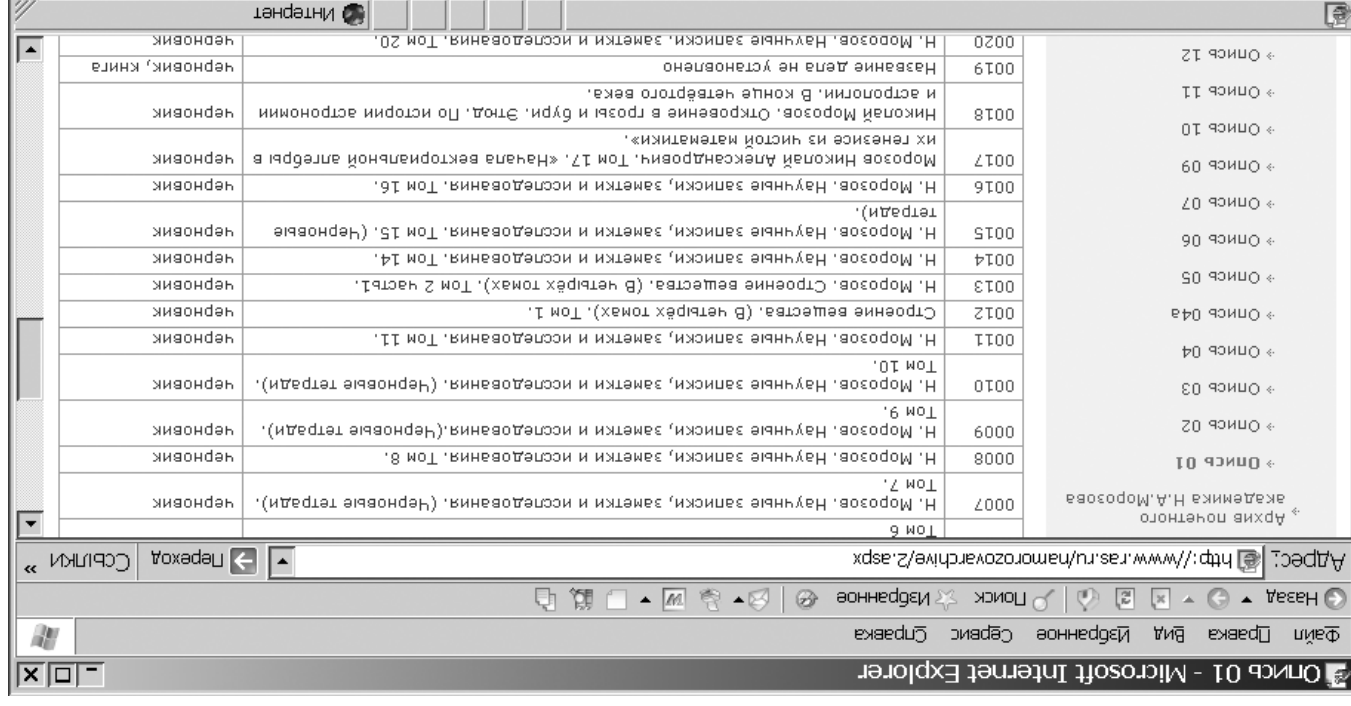


Рис. 1. Интерфейс архива — опись

5. Интеграция архива в портал РАН

Информационный web-портал РАН [1–6] изначально создавался с целью объединения различных информационных ресурсов, так или иначе связанных с деятельностью Академии. Электронный архив «Научное наследие» для этого портала является внешней системой (информационным источником), для интеграции которой в портал следует использовать технологию информационных компонентов [12]. Для этого необходимо было разработать собственный информационный компонент архива, обеспечивающий следующую функциональность:

- подключение к структуре отображения сайта портала элементов структуризации архива (фонды и описи);
- подключение к системе управления содержанием портала шаблонов, обеспечивающих отображение данных архива;
- подключение к серверу интеграции и доступа портала интерфейсов доступа к данным архива (адаптер архива);
- интеграцию с подсистемой безопасности портала.

5.1. Подключение к структуре отображения сайта. Подсистема представления портала [2, 3] поддерживает ряд инструментов, обеспечивающих возможности встраивания в структуру отображения сайта портала данных внешних систем. Использование этих инструментов (разделы, меню, элементы оформления и т. п.) необходимо для предоставления пользователям удобного, единообразного и непротиворечивого интерфейса для доступа к различным интегрированным в портал системам. В случае электронного архива «Научное наследие» ключевой задачей было интегрировать в подсистему представления портала структуру архива. Функциональность портала позволяет подключить структуру к основной структуре отображения и таким образом построить динамическое представление для информационных источников. После внедрения такой структуры постоянная интерфейсная часть отображения портала (разделы, меню и т. п.) генерируются автоматически. Для этих целей в подсистеме представления определён программный интерфейс и схема XML, посредством которых импортируется структура.

Поскольку структура архива сама по себе является иерархической (фонд–опись–дело), задача заключается в формировании древовидной XML-структуры в терминах портала и реализации

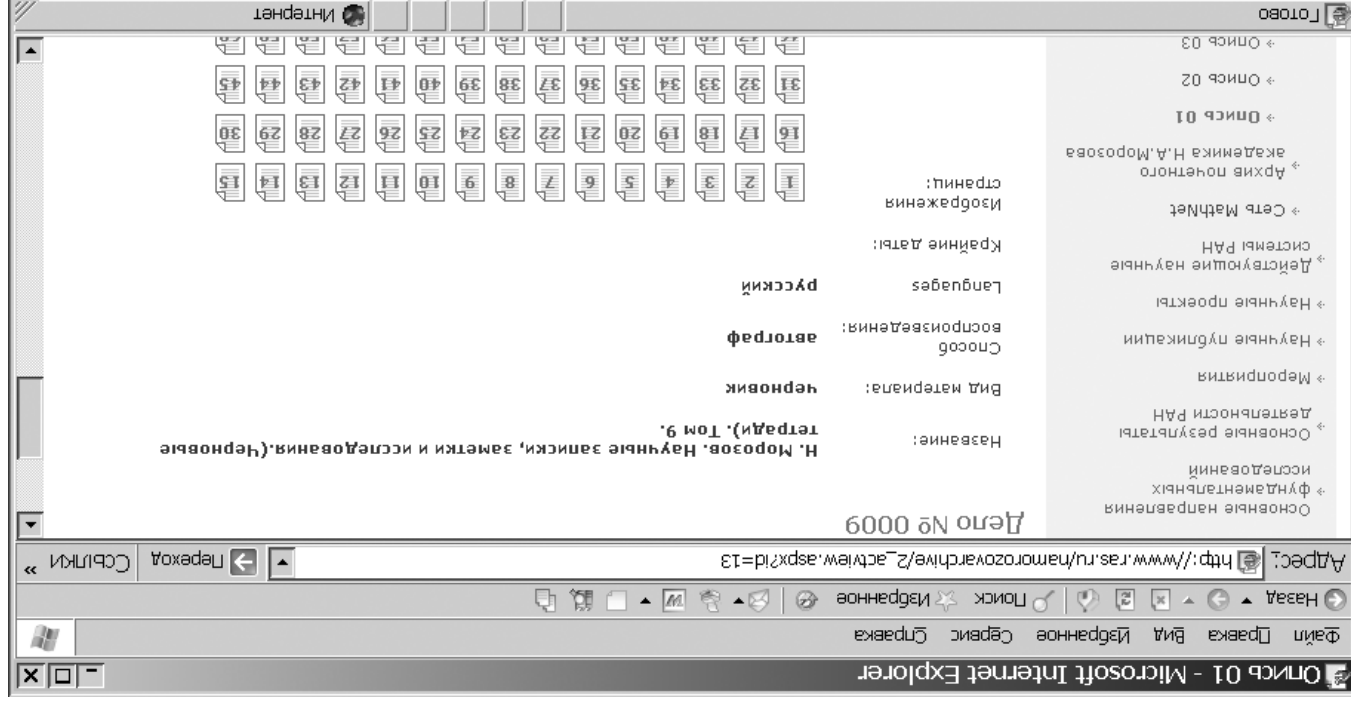


Рис. 2. Интерфейс архива – дело

предопределённого программного интерфейса для предоставления этой структуры.

Первой задачей является получение данных, на основе которых можно построить XML. Сами данные находятся в базе данных под управлением Microsoft SQL Server. Для получения данных существует множество путей. Несомненно, самый простой из них — работа с СУБД через ODBC или OLE DB-интерфейс, но такой способ является не вполне гибким для решения задач интеграции систем при условии их возможной распределённости. Наличие самостоятельного web-сайта, код которого уже содержит нужные процедуры работы с базой данных, определило использование web-сервиса в качестве механизма доступа к данным.

Функционал web-сайта электронного архива «Научное наследие» был дополнен возможностями web-сервиса: определены и реализованы необходимые функции. Такое решение, хотя и увеличивает накладные расходы при работе с данными, но позволяет обеспечить доступность информации с использованием только протокола HTTP, что, соответственно, существенно повышает гибкость системы.

Полный текст WSDL-описания web-сервиса доступен по адресу <http://www.ras.ru/marchive/asvc.asmx?WSDL>.

Построение и интеграция субструктуры выполняется специальным модулем, именуемым информационным компонентом [12]. Этот компонент через web-сервис получает данные из базы данных архива и формирует субструктуру в терминах портала.

5.2. Подключение к системе управления содержанием.

Система управления содержанием в Информационном портале РАН [2, 3] построена таким образом, что позволяет несколькими способами формировать информацию из внешних источников для представления в портале.

Одна из функций этой системы состоит в поддержке (подключении, настройке параметров, включении в структуру отображения и пр.) шаблонов, обеспечивающих отображение данных систем на страницах сайта портала. Разработка таких шаблонов в интересах архива не представляла особого труда, так как собственные ASPX-страницы сайта архива легко модернизировались под требования портала. Более того, подготовить шаблоны представления для архива даже легче, поскольку значительную часть

подготовки пользовательского интерфейса система управления содержанием (рис. 3) берёт на себя: автоматически создаются стандартные для портала элементы дизайна, размещаются управ-

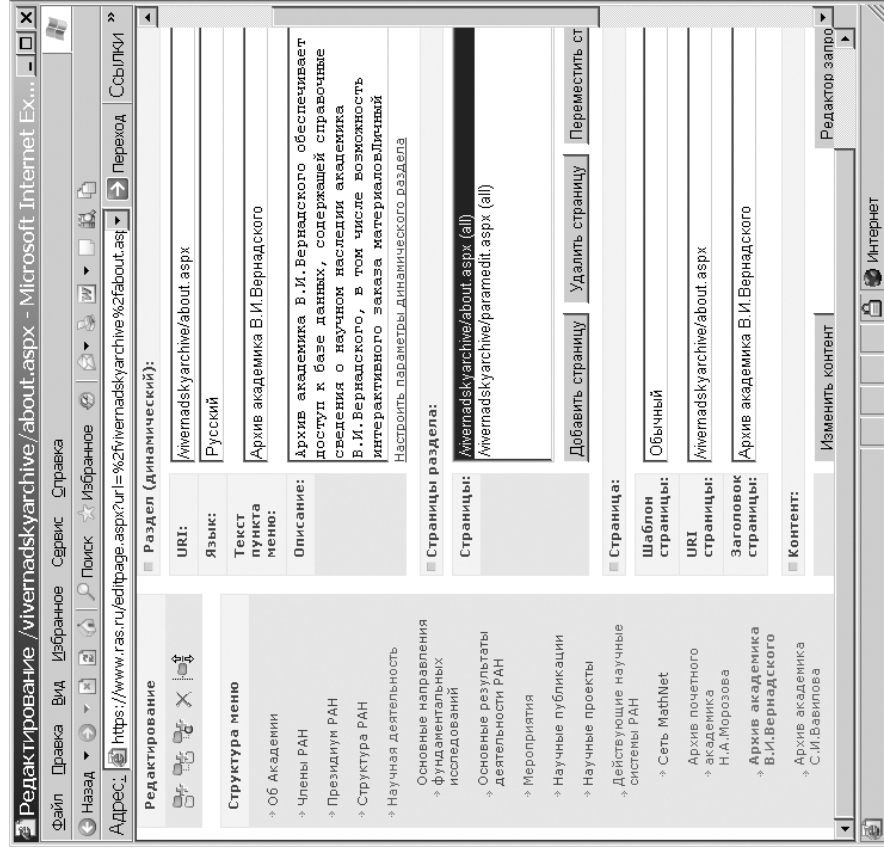


Рис. 3. Интерфейс системы управления содержанием

ляющие компоненты (поиск, регистрация и т. п.), в шаблон включается универсальный компонент доступа к данным, наполнение которого обеспечивает адаптер архива — модуль сопряжения с источником.

Подключиться к системе управления содержанием можно и другим способом, а именно, путём использования готового универсального шаблона, содержательная часть которого также предоставляется адаптером. Когда в портале строится отобра-

жение страницы на основе шаблона, то в момент построения содержательной части происходит вызов соответствующего адаптера, который экспортирует помимо данных ещё и компонент отображения, с помощью которого и происходит отображение содержательной части.

При подключении архива к порталу были использованы оба способа.

5.3. Подключение к серверу интеграции и доступа. Сервер интеграции и доступа портала [2, 3] поддерживает механизм передачи данных от внешней системы, подключённой к порталу, к стандартным компонентам доступа к данным, автоматически включаемым в шаблоны отображения. Данный механизм может основываться на разных протоколах. В текущей реализации поддерживается доступ (имеются адаптеры) к хранилищам ADO.NET, ODBC и технология Web-Services. Для интеграции архивной информации в портал использовалась технология web-сервисов. При обращении пользователя к соответствующему шаблону подсистема представления портала обращается к серверу интеграции и доступа, который передаёт указанный запрос соответствующему адаптеру. Адаптер, в свою очередь, получает данные и передаёт их компоненту доступа к данным шаблона. Отобразить полученные таким образом данные — техническая задача, решаемая шаблоном архива, интегрированным в систему управления содержанием портала.

Подключение информационного источника, в том числе электронного архива, требует определения некоторых параметров. Механизмы портала позволяют внедрять административный интерфейс управления источником в систему управления содержанием: если настраиваются шаблоны представления информации данных внешнего информационного источника, то при настройке параметров производится импорт административного интерфейса через информационный компонент. Для настройки параметров информационного источника, администратор портала имеет возможность использовать этот административный интерфейс, который содержит все управляющие элементы для задания параметров, нужных для взаимодействия с внешней системой.

5.4. Интеграция с системой безопасности портала. Подключённый к порталу шаблон автоматически получает функциональность, необходимую для использования механизмов защиты, встроенных в портал, в частности сервиса аутентификации [13].

Ограниченный доступ, таким образом, может быть реализован для любых шаблонов (а значит, и данных) фондов архива. На текущий момент все документы архива находятся в открытом доступе, но при необходимости он может предоставляться только авторизованным пользователям. Причём активизировать любые ограничения можно с использованием обычного интерфейса администратора безопасности портала [13] без какой-либо дополнительной доработки шаблонов архива.

6. Заключение

В настоящей статье описана одна из информационных систем, созданных и поддерживаемых организациями РАН и представляющих несомненный интерес для широкого круга научной общественности. Для представления этих систем всем заинтересованным категориям пользователей в качестве компетентного общедоступного источника и создавался Информационный web-портал РАН www.ras.ru. На примере электронного архива «Научное наследие» продемонстрирована технология создания электронного представления для такого рода ресурсов, а также возможность и лёгкость их интеграции в среду информационного портала, обеспечивающего знакомство с этими ресурсами широкой научной общественности.

Список литературы

1. Соколов И. А., Босов А. В., Бездудинский А. Н. О Информационном web-портале Российской академии наук // Системы и средства информатики. Вып. 13. — М.: Наука, 2003. — С. 119–138.
2. Босов А. В., Иванов А. В. О реализации системы управления содержанием информационного web-портала // Информационные технологии и вычислительные системы. № 4. — М.: ИМВС РАН, 2004. — С. 85–103.
3. Босов А. В., Иванов А. В. Технология управления содержанием информационного портале РАН // Системы и средства информатики. Вып. 15. — М.: Наука, 2005. — С. 260–283.
4. Босов А. В., Иванов А. В., Полушкин А. Н., Чавтараев Р. Б. Управление сайтом информационного web-портала // Системы и средства информатики. Вып. 15. — М.: Наука, 2005. — С. 233–259.

5. Россия. Веб-портал РАН обеспечивает эффективный доступ научного сообщества к актуальной информации // Информационный бюллетень Microsoft. 2004. Вып. 26. С. 67–68.
6. Российская академия наук реализует систему управления содержанием своего информационного веб-портала на основе платформы Microsoft.Net. Примеры внедрения — <http://www.microsoft.com/rus/Casestudies/CaseStudy.aspx?id=318>.
7. Об архивном фонде Российской Федерации и архивах. Российская Федерация. Основы законодательства от 7 июля 1993 года № 5341-1 // Архивное дело / Под. ред. А. С. Кольянова. — М.: Книга сервис, 2003.
8. ГОСТ Р. 51141-98. Делопроизводство и архивное дело. Термины и определения // Архивное дело / Под. ред. А. С. Кольянова. — М.: Книга сервис, 2003.
9. Об утверждении временного порядка автоматизированного государственного учёта документов архивного фонда Российской Федерации, хранящихся в государственных и муниципальных архивах. Приказ Федеральной архивной службы России от 23 октября 2000 года № 64 // Архивное дело / Под. ред. А. С. Кольянова. — М.: Книга сервис, 2003.
10. Основные правила работы ведомственных архивов. Приказ Главархива СССР от 5 сентября 1985 года № 263 // Архивное дело / Под. ред. А. С. Кольянова. — М.: Книга сервис, 2003.
11. Основные правила работы архивов организаций. Решение Коллегии Росархива от 6 февраля 2002 года // Архивное дело / Под. ред. А. С. Кольянова. — М.: Книга сервис, 2003.
12. Босов А. В., Чавтараев Р. Б. Управление информационными компонентами веб-портала РАН // Системы и средства информатики. Вып. 13. — М.: Наука, 2003. — С. 156–171.
13. Босов А. В., Полушкин А. Н. О реализации сервиса аутентификации веб-портала // Информационные технологии и вычислительные системы. № 3. — М.: ИМВС РАН, 2005. — С. 50–60.