

На правах рукописи

КОВЯЗИНА ЕЛЕНА ВАСИЛЬЕВНА

**ТЕХНОЛОГИЯ ДОСТУПА К ДОКУМЕНТАМ
В НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ ОРГАНИЗАЦИИ**

Специальность 05.25.05 – Информационные системы и процессы,
правовые аспекты информатики

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата технических наук

Новосибирск
2007

Работа выполнена в Институте вычислительного моделирования
Сибирского отделения РАН и Институте вычислительных технологий
Сибирского отделения РАН.

Научный руководитель член-корреспондент РАН, профессор
Федотов Анатолий Михайлович

Официальные оппоненты: доктор технических наук, профессор
Елепов Борис Степанович

кандидат технических наук
Мазов Николай Алексеевич

Ведущая организация: ФГОУ ВПО «Сибирский
федеральный университет»

Защита состоится «9» ноября 2007 г. в 17.45 на заседании
Диссертационного совета Д003.046.01 при Институте вычислительных
технологий Сибирского отделения РАН по адресу: пр. Академика
М. А. Лаврентьева, д. 6, 630090, Новосибирск, Россия. Тел.: (383)
3331882, факс: (383) 3306342. E-mail: dsovet@ict.nsc.ru

С диссертацией можно ознакомиться в специализированном
читальном зале вычислительной математики и информатики ГПНТБ
СО РАН.

Автореферат разослан « 8 » октября 2007 г.

Ученый секретарь
диссертационного совета,
доктор физико-математических
наук, профессор

Чубаров Л. Б.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования. Определяющим в государственной информационной политике России последних лет является переход к информационному обществу. Актуальными проблемами на этом пути являются проблемы развития технологического базиса информационного общества, существенными компонентами которого являются:

- национальные информационные ресурсы, значительная часть которых сосредоточена в библиотеках, и обеспечение широкого доступа к ним;
- информационно-коммуникационная инфраструктура, представленная территориально распределёнными государственными и корпоративными сетями;
- информационные, компьютерные и телекоммуникационные технологии, системы и средства их реализации, сетевые технологии обеспечения доступа к информации.

Развитие средств автоматизации в традиционных библиотеках создало условия для синтеза этих составляющих и превращения традиционной библиотеки в электронную или гибридную, где сосуществуют традиционные и электронные технологии.

Несмотря на то, что термину «электронная библиотека» в разных источниках даются разные определения, все они содержат, в том или ином контексте, понятие «доступ к информационным ресурсам». Согласно одному из определений, «электронная библиотека – это созданная на профессиональной основе тематически ориентированная (или организованная иным образом) система доступа к удаленным или локальным электронным ресурсам, способная обслуживать электронными ресурсами и специализированной помощью локальных и удаленных пользователей»¹. Таким образом, разработка технологии доступа к документам, как определяющей составной части электронной библиотеки научно-исследовательской организации, является **актуальной задачей**.

Степень изученности проблемы и методология исследования. В 1960-е гг. Библиотекой Конгресса США был разработан машиночитаемый формат MARC, используемый для корпоративной каталогизации и обмена информацией между библиотеками, а в 70-х гг. – сетевой протокол Z39.50, предназначенный для связи между компьютерами и обмена информацией между ними. Два этих факта явились определяющими в дальнейшем развитии технологий доступа к документам в библиотеках. Семейство форматов MARC использовалось совместно с Z39.50 для решения задач

¹ Земсков А. И. Электронные библиотеки [Текст] / А. И. Земсков, Я. Л. Шрайберг. – М. : Либерея, 2003. – 352 с.

поиска и онлайн-доступа к документам, а также электронной доставки копий документов. Доступ к документам и их доставка – противоположно направленные процессы, в которых доставка предполагает пассивную роль пользователя, характерную для технологий работы крупной библиотеки. Стремление доставить документ пользователю в привычном для него виде определило использование для этих целей факсов, а впоследствии электронной почты. В том же направлении развивались и появившиеся позже технологии электронных издательств, предоставлявших свои ресурсы посредством телекоммуникационных сетей. Появление и быстрое развитие Web-среды определило приоритетное развитие технологий онлайн-доступа. Однако Web предполагает работу с неорганизованными нестабильными ресурсами, требующими значительных усилий для их семантического описания, развитых поисковых средств и информационных сервисов.

В научной организации пользователи нуждаются в совместном использовании как ресурсов Интернет, так и традиционных ресурсов библиотек. К тому же, высокий уровень их квалификации предполагает большую самостоятельность в процессе доступа к ресурсам и, как следствие, большую требовательность к их упорядоченности и систематизации. Системы доступа к документам российских разработчиков, такие как АБИС «РУСЛАН» («Открытые библиотечные системы», СПбГТУ), библиотечный Интернет-комплекс (ГПНТБ России), рассчитаны на технологии работы крупных библиотек, обладают значительной функциональной избыточностью и требуют дополнительных усилий для обеспечения интеграции ресурсов различных производителей.

Методологической основой исследования явились положения изложенные в трудах в области информатики, в частности, в области проектирования и разработки информационных систем Дж. Солтона, Дж. Мартина, А. Б. Антопольского, Ф. С. Воройского; А. М. Федотова, Н. Е. Каленова, Б. С. Елепова, С. Р. Баженова; компонентному и функциональному составу электронных библиотек Я. Л. Шрайберга, А. И. Земскова, А. И. Бродовского, В. Армса, Р. Мэррея; проектированию распределенных систем О. Л. Жижимова, Н. А. Мазова, А. И. Племнека; перспективным технологиям Интернет М. Р. Коголовского; системам электронной доставки документов В. А. Глухова; теории систем и методологии системных знаний М. Месаровича, Н. Н. Моисеева и лингвистическому обеспечению информационных систем Э. Р. Сукиасяна.

Объектом исследования в данной работе являются системы доступа к информационным ресурсам в распределенных хранилищах.

Предмет исследования: компонентный состав систем доступа к информационным ресурсам и информационные потоки внутри каждого компонента, модели информационных ресурсов в системах доступа.

Цель диссертационной работы состоит в выработке специфичных для научно-исследовательской организации технологических решений обеспечения доступа к документам, которые допускают как унификацию доступа к разнородным информационным ресурсам, так и интеграцию с другими информационными системами.

Достижение цели потребовало решения следующих задач:

1. На основе анализа информационных потоков в библиотеке научно-исследовательской организации (НИО) разработать модель информационных процессов технологии доступа к документам

2. Определить модель хранения данных и схемы описательных метаданных документов библиотеки

3. Обосновать необходимость выбора определенных программных средств и реализовать дополнительные программные инструменты, обеспечивающие выполнение технологических требований к доступу к документам.

4. На основе разработанной технологии реализовать доступ к документам в сети библиотек Красноярского научного центра (КНЦ) СО РАН.

Методы исследования базируются на теории сложных систем, математической теории множеств, теории информации и информационных процессов, основах компьютерных и телекоммуникационных технологий, а также библиотечных и библиографических стандартах и нормативных документах Интернет.

Научная новизна связана с моделированием информационных процессов в технологии доступа к документам, основанной на специфике научно-исследовательской организации, а также разработкой модели распределенного хранения электронных документов.

Практическая ценность. Разработка и реализация технологии и модели хранения данных выполнена как составная часть работы по проектам РФФИ № 98-07-90128-в «Красноярская информационная сеть научных институтов и вузов» в 1998–1999 гг., № 00-07-90340-в «Создание интегрированной сети информационного сетевого центра в г.Красноярске» в 2000–2001 гг., № 02-07-90135-в «Создание Красноярской сети параллельных вычислений» в 2002–2004 гг., № 05-07-90201-в «Разработка Красноярской городской информационно-вычислительной сети науки и высшей школы» в 2005–2007 гг. Предложенная технология доступа к документам используется в информационной сети библиотек Красноярского научного центра (КНЦ) СО РАН. Технологические требования к доступу позволяют адаптировать

разработанную технологию в систему библиотек иной ведомственной или организационной принадлежности, что показал опыт успешного внедрения технологии в библиотеке Политехнического института Сибирского федерального университета.

Достоверность и обоснованность результатов диссертационной работы подтверждаются успешным практическим использованием технологии в сети библиотек КНЦ СО РАН, а также:

1. Результатами анализа существующих технологий, протоколов и стандартов в области построения систем доступа к документам из Интернет.

2. Применением в функциональных модулях технологии программных средств, основанных на международных стандартах.

3. Применением технологической модели для реализации доступа к разнородным информационным ресурсам с различными правами доступа.

Основные положения, выносимые на защиту:

1. Технология доступа к документам, основанная на специфике работы научно-исследовательской организации, должна отвечать следующим основным требованиям: интероперабельность, базирующаяся на стандартах открытых систем, интегрируемость в единую информационную среду РАН, распределенность документов по местам хранения и децентрализация администрирования коллекций документов, использование готовых программных решений, обеспечение отложенного доступа системой доставки документов.

2. Для обеспечения эффективной работы системы, построенной на программных решениях различных разработчиков, технологию доступа к документам целесообразно строить как систему взаимосвязанных функциональных модулей и информационных хранилищ.

3. Модель документов в хранилище данных строится на раздельном хранении текста документа и его описательных метаданных, основанных на международных стандартах и рекомендациях. Такой тип хранения позволяет вести распределенный поиск и хранить электронные документы по месту их производства.

4. Разработанная технология позволяет организовать доступ к различным документам, от печатных до электронных, с различными правами доступа к ним, что иллюстрирует реализация следующих баз данных:

- Аннотированный каталог книг и продолжающихся изданий библиотеки Института вычислительного моделирования (ИВМ) СО РАН;
- Архив научных публикаций ИВМ СО РАН;
- Электронные документы по информационным технологиям, математике и механике ИВМ СО РАН.

Личный вклад автора. Основные результаты, представленные в работе, получены непосредственно автором, а именно: технологические требования,

предъявляемые к доступу; модель информационных процессов в технологии доступа; модель информационного хранилища; базы данных «Архив научных публикаций ИВМ СО РАН», «Электронные документы по информационным технологиям, математике и механике ИВМ СО РАН» и «Аннотированный каталог книг и продолжающихся изданий ИВМ СО РАН»

Апробация работы. Научные результаты, изложенные в диссертации, докладывались на следующих конференциях: VII, IX, X, XI, XIII, XIV международных конференциях «Библиотеки и ассоциации в современном мире науки, культуры, образования и бизнеса (Крым)» (г. Судак, Украина) в 2000, 2002–2005, 2006–2007 гг., международных конференциях «ЛИБКОВ» (г. Москва) в 2004–2006 гг., VII и IX международных конференциях по электронным публикациям «EI-Pub» (г. Новосибирск) в 2002 и 2004 гг., региональных научно-практических конференциях «Научные библиотеки в новом тысячелетии: проблемы взаимодействия ресурсов» (Иркутск, 2001 г.), «Библиотека и духовная культура нации» (г. Новосибирск) в 2004 г., «Электронные ресурсы региона» (г. Новосибирск) в 2004 г., «Библиотеки на пути к открытому информационному пространству» (г. Красноярск) в 2002 г.

Публикации. По теме диссертации опубликовано 14 научных работ, в т. ч. 2 публикации – в изданиях, рекомендованных Высшей аттестационной комиссией.

Структура и объем работы. Диссертация состоит из введения, 4 глав, заключения, списка использованной литературы, включающего 149 названий, 5 приложений. Основное содержание работы изложено на 147 страницах текста, общее количество страниц – 234. Работа проиллюстрирована 10 рисунками.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, рассмотрена изученность проблемы, представлены цель и задачи исследования. Определена научная новизна и практическая значимость, приведены основные результаты работы. Сформулированы основные положения, выносимые на защиту.

Первая глава «Анализ технологий доступа к документам» посвящена обзору истории и современного состояния систем доступа к документам и их анализу.

В разделе 1.1 приведен обзор зарубежных проектов доступа и доставки документов. Выделены основные этапы и направления развития технологий доступа к документам: разработка базовых стандартов, формирование и накопление электронных ресурсов, развитие поисковых систем Интернет и систем доставки документов.

В разделе 1.2 приведен обзор отечественных разработок технологий доступа к документам, определены их особенности и характерные черты.

В разделе 1.3 приведен анализ имеющихся программных решений в области доступа к документам. В результате анализа выделены обязательные составные части технологии доступа к документам, а именно:

1. Описательные метаданные в формате семейства MARC, или ином стандартизованном формате.

2. Сетевые протоколы, позволяющие унифицировать доступ к массивам метаданных, сформированным различными системами и установленным на различных аппаратных платформах, а также оперировать распределенными данными.

3. Унифицированная поисковая среда и форматы хранения документов.

Остальные возможные требования к технологиям доступа можно считать специальными и определяемыми в зависимости от круга пользователей, для которых технология предназначена.

Все реализованные технологии доступа к ресурсам в России можно условно разбить на четыре группы, в зависимости от характера информационных ресурсов и требуемой функциональности:

1. Системы доступа к документам в крупных библиотеках реализуют доступ к централизованно хранящимся данным. Как правило, в этой группе не используется одновременный поиск по нескольким базам данных. Системы ориентированы исключительно на используемую в конкретной библиотеке автоматизированную библиотечно-информационную систему (АБИС), а также рассчитаны на технологии и формы отчетности, принятые в таких организациях.

2. Большинство российских корпоративных библиотечных систем, работающих с распределенными данными, не обеспечивают доступ к полным текстам и систему заказа электронных копий документов, а также не используют поиск по словарям метаданных.

3. Системы доступа издательств и других поставщиков информационных ресурсов осуществляют коммерческий доступ к документам, преимущественно к журнальным статьям, поэтому системы используют закрытые протоколы, ориентированы на лицензирование доступа и жесткое разграничение пользователей по правам использования ресурсов.

4. Поисковые машины Интернет имеют дело с большим количеством хаотичных, тематически неорганизованных ресурсов, а также с усредненным пользователем с неопределенным кругом интересов.

Научно-исследовательская организация имеет свою специфику и контингент пользователей, поэтому не может прямо заимствовать ни одну из описанных технологий. Для организации доступа к документам в такой

организации необходим специальный подход, основанный на потребностях пользователей и доступных организации информационных ресурсах.

Вторая глава «Модель доступа к документам в научно-исследовательской организации» посвящена разработке информационной модели технологии доступа к документам.

В разделе 2.1 приведено определение доступа, компонентный состав технологии доступа к документам, специфика научно-исследовательской организации, определяющая особенности технологических подходов. Обозначены задачи, которые необходимо решить для обеспечения доступа к документам. На основании пользовательских рекомендаций определены технологические требования к доступу:

1. Открытость и расширяемость. Доступ к документам из Интернет базируется на стандартах открытых систем, что обеспечивает переносимость технологических блоков на иные аппаратные платформы, а также подключение источников информации других производителей. Обеспечивается использованием Web-технологий, коммуникативного формата RUSMARC и сетевых протоколов Z39.50.

2. Простота управления. Достигается использованием в технологии функциональных модулей, определяемых общностью информационных процессов и используемыми информационными технологиями. Разбиение на модули позволяет использовать в них готовые программные решения разных разработчиков. Модульность позволяет автономно обслуживать каждый блок, а также модернизировать и заменять используемые в них программные средства по мере развития соответствующих информационных технологий.

3. Централизованный пользовательский интерфейс. Поиск документа и доступ к нему производится из «единого окна». Для пользователя система выглядит централизованной, так как её распределенность скрыта интерфейсом.

4. Распределенность документов и коллекций по местам хранения и децентрализация их администрирования. Документы, коллекции и группы коллекций хранятся там, где образуются.

5. Интегрируемость на уровне баз данных. Технология обеспечивает доступ к ресурсам различных производителей, а также допускает логическую группировку коллекций документов, в том числе созданных с помощью различных СУБД, с целью формирования тематических ресурсов и ресурсов, скомпонованных по региональному признаку.

6. Дифференциация прав доступа. Для обеспечения условий лицензионных соглашений и авторских прав уровень доступа к базам данных организации и отдельным документам определяется индивидуальными правами пользователя или группы пользователей.

7. Унифицированные описательные метаданные и обеспечение многофункционального поиска по ним. Для обеспечения доступа к документам с помощью различных поисковых интерфейсов документы различного типа и вида представлены в системе унифицированным способом описания двух типов, допускающих взаимную конверсию:

а. В блоке формирования метаданных документ описывается АБИС в формате семейства MARC;

б. Документ, поступающий в хранилище документов через экспертное хранилище, получает RDF-описание в схеме Dublin Core (DC) или MODS.

8. Унифицированные форматы хранения документов. Формат хранения документа должен допускать его семантический и синтаксический анализ для последующего поиска по полному тексту.

9. Вариативность форматов вывода на экран результатов поиска документов – от MARC-форматов до формата каталожной карточки. Доступ к электронному документу on-line инкапсулирован в формате вывода как URL-ссылка.

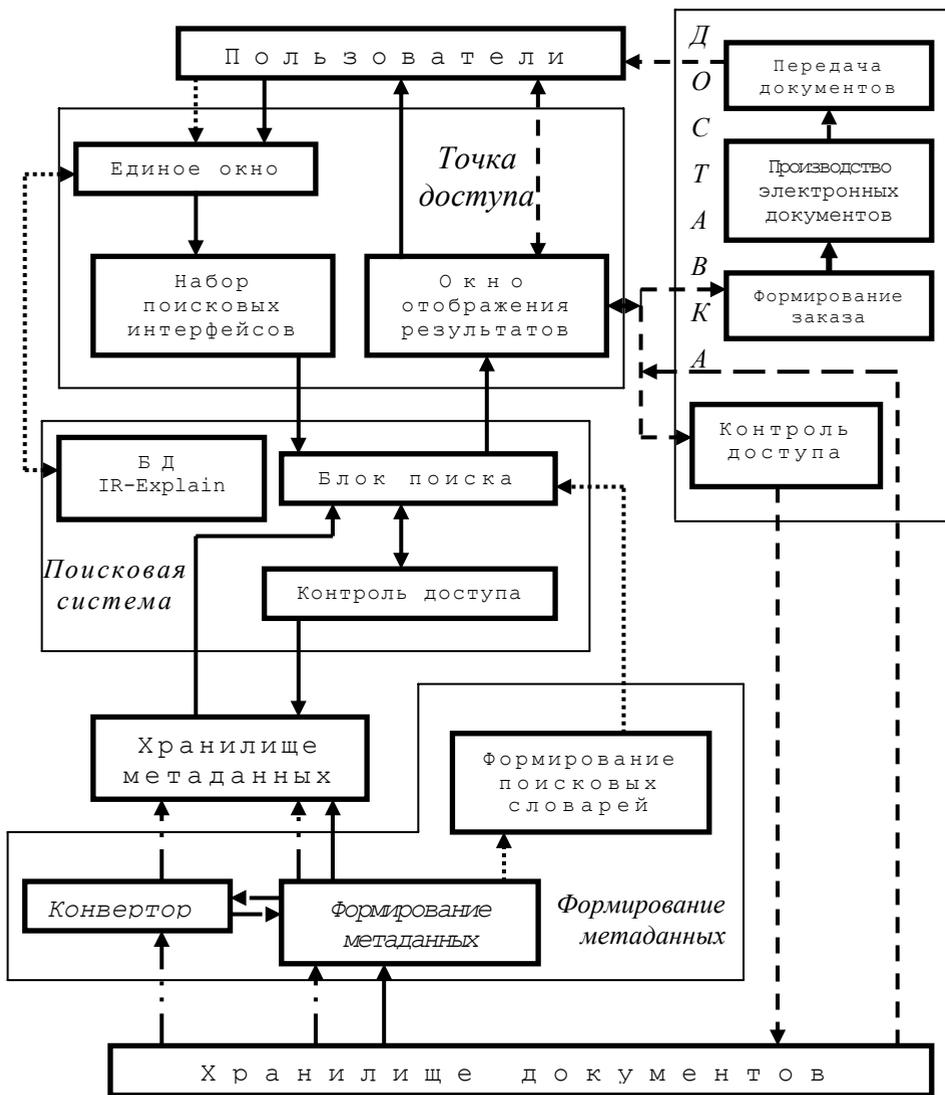


Рис. 1. Модель информационных процессов технологии доступа

10. Возможность «отложенного доступа», т. е. регистрация заявки, конверсия документа в электронный вид с сохранением и последующей доставкой пользователю через систему доступа или электронной почтой.

В разделе 2.2 приведена общая схема информационных процессов технологии доступа к документам (рис. 1).

В технологии выделены функциональные модули, реализующие отдельные информационные процессы, и модули информационных ресурсов технологии - хранилища данных.

Функциональный модуль – это технологический блок целевого назначения с собственными информационными потоками.

Выделение функциональных модулей обусловлено следующими признаками их внутреннего единства:

1. Однородностью используемых в пределах модуля информационных технологий и программных средств их реализации.

2. Возможностью обслуживания и модернизации каждого модуля специалистом, имеющим навыки работы только с конкретными программными продуктами.

Для хранения информационных ресурсов технологии используются четыре хранилища:

1. Распределенное хранилище документов.

2. Распределенное хранилище метаданных документов.

3. База данных описаний коллекций.

4. Временное экспертное хранилище документов перед помещением их на постоянное хранение.

Раздел 2.3 посвящен описанию функциональных модулей технологии доступа. Выделены четыре модуля: единая точка доступа, поисковая система, формирование метаданных и доставка.

Единая точка доступа реализует технологию «единого окна» и предназначена служить пользовательским интерфейсом доступа к документам. Разработка и реализация точки доступа к документам осуществляются на основе технологий Web. С помощью точки доступа пользователь извещается о доступных ему базах данных и может самостоятельно выделить список ресурсов для поиска необходимого ему документа. Точка доступа оперирует тремя информационными потоками:

1) Поисковые запросы пользователей с вектором «пользователь - единое окно - набор поисковых интерфейсов – поисковая система – окно отображения результатов – пользователь».

2) Информация о массивах метаданных документов с вектором «пользователь – единое окно - поисковая система – единое окно – пользователь».

3) Запросы на доставку документов с вектором «пользователь – окно отображения результатов – доставка – окно отображения результатов – пользователь».

Поисковая система предназначена для поиска документов по набору их поисковых индексов. Базовым требованием к поисковой системе является возможность распределенного поиска по массивам метаданных, размещенным в Интернете. Для этой цели используется сетевой протокол Z39.50 (стандарт ISO 23950). Метаданные документов группируются в массивы, определяемые видовыми и/или тематическими характеристиками документов. Описания массивов хранятся в базе данных IR-Explain. Доступ пользователей к массивам определяется правами доступа и обеспечивается системой контроля, встроенной в протокол Z39.50. В поисковой системе определены три информационных потока:

1) Поисковые запросы пользователей с вектором «точка доступа – блок поиска – контроль доступа – хранилище метаданных - блок поиска – точка доступа», либо в случае отказа в доступе «точка доступа – блок поиска – контроль доступа – блок поиска – точка доступа».

2) Информация о массивах метаданных с вектором «точка доступа – БД IR-Explain – точка доступа».

3) Лингвистическое обеспечение поиска с вектором «формирование метаданных – блок поиска».

Модуль формирования метаданных предназначен для обработки полных текстов документов перед размещением их в хранилище. В процессе обработки документа блоком формирования метаданных создается вторичный документ - библиографическое описание. Представление библиографического описания, а также результатов индексирования и реферирования документа в стандартной машиночитаемой форме семейства MARC создает библиографическую запись или поисковый образ документа, который может быть использован поисковой системой. Библиографические записи компонуются в именованные массивы записей, индексируются и размещаются в хранилище метаданных. Для этого пригодна автоматизированная библиотечно-информационная система (АБИС), используемая в библиотеке организации. Однако для совместного функционирования модуля с поисковой системой используемый технологией сервер Z39.50 должен быть укомплектован соответствующим провайдером данных, отображающим данные АБИС в абстрактную схему данных. Модуль содержит три функциональных блока и обслуживает следующие потоки информации:

1) Формирование описательных метаданных в формате MARC с вектором «хранилище документов – формирование метаданных – хранилище метаданных»

2) Формирование метаданных в формате DC (или MODS) с вектором «хранилище документов - формирование метаданных – конвертор – хранилище метаданных». Возможен и встречный поток для преобразования метаданных в формате DC, содержащихся в электронном документе, в формат MARC с последующей их корректировкой, с вектором «хранилище документов – конвертор – формирование метаданных – хранилище метаданных».

3) Лингвистическое обеспечение поиска с вектором «формирование метаданных – формирование поисковых словарей - поисковая система».

Модуль доставки функционирует в зависимости от способа хранения требуемого документа и прав доступа к нему:

1. Если электронная версия документа находится в открытом доступе, то модуль производит онлайн-доставку его на компьютер пользователя по URL-ссылке.

2. Если доступ к электронной версии ограничен, пользователь может оформить заказ на доставку и, если возможно, получить его по электронной почте.

3. Если документ существует в печатном виде и возможна его конверсия в электронный вид, регистрируется заказ на доставку, а полученный в результате конверсии документ отсылается пользователю электронной почтой.

В модуле определены следующие информационные потоки:

1) Доставка электронного документа с вектором «точка доступа – контроль доступа – хранилище документов – точка доступа».

2) Доставка электронного документа с ограничениями в доступе с вектором «точка доступа – контроль доступа – точка доступа – формирование заказа – пользователь».

3) Доставка печатного документа, пригодного для оцифровки с вектором «точка доступа – формирование заказа – производство электронного документа – пользователь».

Третья глава «Метаданные и их связь с документами в хранилище» посвящена описанию информационных ресурсов технологии доступа – документов и метаданных.

В разделе 3.1 описаны информационные ресурсы технологии, определен класс принадлежности документов и приведены необходимые определения. Перечислены основные стандарты метаданных, используемые в технологии.

В разделе 3.2 описана модель документа в хранилище. Под документом понимается:

$$d_i = \langle p_i, m_i \rangle,$$

где p_i – содержательная часть (или файл) документа, m_i – метаданные документа. Эквивалентные множества $P=\{p_1, p_2, \dots, p_n\}$ и $M=\{m_1, m_2, \dots, m_n\}$, определяющие массив документов $D=\{d_1, d_2, \dots, d_n\}$, образуют два связанных хранилища данных: хранилище документов (или их полных текстов) и хранилище метаданных. Далее обоснована возможность применения объектной модели документа к его метаданным в технологии доступа к документам. Приведены определения.

Метаданными документа t_i называется пара

$$m_i = \langle S_i, V_i \rangle,$$

где S_i – структура метаданных, V_i – содержание (информационное наполнение) структуры. Структура документа определяется одной из выбранных схем данных. Для публикаций это MARC, DC или MODS. Описательные метаданные (содержание) описывают семантику документа. Описательные метаданные могут являться частью электронного документа или храниться отдельно от него. Для того, чтобы избежать дублирования документов, определен признак идентичности: два документа d_1 и d_2 назовем *экземплярами документа*, если они имеют идентичные метаданные $m_1 = m_2$.

Коллекцией называется множество документов (или их метаданных) с выделенной фиксированной структурой, содержание которых имеет одинаковую тематическую направленность.

Из определения документа понятно, что для множества документов D в хранилище документов понятие коллекции размыто, так как в определении документа не входит в явном виде описание его структуры. Понятие коллекции определено на множестве метаданных. В силу связанности и эквивалентности множеств D и M каждому входящему в коллекцию элементу t_i множества M соответствует элемент d_i множества D .

Итак, все множество метаданных в хранилище метаданных разделено на коллекции. Иными словами коллекция K_j есть подмножество множества M : $K_j \subseteq M$. В общем случае, все множество M является коллекцией изданий, доступных сотрудникам научно-исследовательской организации. Все множество коллекций распределено по местам их формирования в сети. Коллекции могут логически объединяться по какому-либо признаку, например, по виду или характеру документов, организации формирования или географического расположения коллекции. Любая из коллекций K_j может представлять собой распределенный информационный ресурс. Семейство множеств K_j является разбиением множества M , так как $M = \bigcup_{j \in J} K_j$ где J – множество индексов и K_j попарно не пересекаются. Таким образом, в хранилище не существует метаданных, не принадлежащих ни одной коллекции.

В разделе 3.3 описываются особенности представления метаданных документов в Интернет и стандартные схемы представления структуры документов: MARC, Dublin Core и MODS.

В четвертой главе «Реализация технологии доступа к документам в Красноярском научном центре СО РАН» описано применение технологии доступа к документам в сети библиотек КНЦ СО РАН и организация на её основе доступа к разнородным коллекциям документов.

В разделе 4.1 представлено описание шлюза Z39.50-HTTP как единой точки доступа к документам. Построено дерево информационных ресурсов, доступных из единого окна. Определены виды опубликованных ресурсов и способы их организации. Описано используемое программное обеспечение и его функциональные характеристики. HTML-код страницы «единого окна доступа» приведен в приложении 1. Приложение 2 содержит запись базы данных IR-Explain, описывающую аннотированный каталог книг библиотеки ИВМ СО РАН.

В разделе 4.2 приведено применение технологии доступа к печатным документам библиотеки организации. Показано, что для отражения семантики печатного документа электронный каталог должен содержать аннотации или рефераты. Описаны методы создания сводных каталогов и преимущества их использования в целях получения информации о печатных документах, хранящихся в различных библиотеках.

В разделе 4.3 представлена база данных научных публикаций сотрудников Института вычислительного моделирования СО РАН, построенная на принципах архива открытого доступа. Каждый документ в архиве снабжен метаданными в схеме DC, которые позволяют дополнительно интегрировать документ в Web-среду. Для преобразования описательных метаданных документа из схемы DC в схему MARC, и наоборот, применяются готовые конверторы, встроенные в АБИС, и программы собственной разработки, одна из которых приведена в приложении 4.

В разделе 4.3 продемонстрировано применение технологии для доступа к коллекции электронных документов Института вычислительного моделирования СО РАН. Коллекцию составляют частные собрания сотрудников Института, а доступ к документам коллекции регламентируется авторским правом. Показано, что заложенная в технологию трехступенчатая система защиты обеспечивает дифференцированный доступ к документу для различных категорий пользователей.

Выводы и рекомендации. Технологии доступа к данным являются интенсивно развивающейся областью информационных технологий. Существует большое количество систем доступа к документам как основанных на поиске в базах данных, так и использующих поиск в Web.

Множество предлагаемых технологий доступа затрудняют выбор наиболее подходящей технологии, настроенной на специфические требования пользователей конкретной организации. Для выбора подходящих методов доступа в работе проведен обзор существующих систем доступа в их историческом развитии. В результате анализа выделены общие требования к технологиям доступа: стандартизованные структуры (форматы или схемы) описания документов, трактуемые как метаданные, и наличие стандартизованных правил (протоколов) для доступа к метаданным по сетям. В обзоре выделены основные особенности различных информационных технологий, используемых для доступа к документам. Определены различия между доступом и доставкой документов.

В работе впервые:

- описана модель информационных процессов в технологии доступа к документам на основе функциональных модулей;
- определен комплекс технологических требований обеспечения доступа к разнородным документам в библиотеке НИО;
- разработана модель распределенного хранения документов, обеспечивающая эффективный поиск и доступ к ним;
- на основе разработанной технологии и модели хранения документов реализована прикладная система доступа к документам библиотеки Красноярского научного центра СО РАН.

Для разработки технологической модели доступа к документам в научно-исследовательской организации выделены основные особенности такой организации, определяющие пользовательские требования к доступу. Выбор Web-среды для доступа к документам, общие требования к технологиям доступа, а также особенности научно-исследовательской организации позволили сформулировать набор требований к технологии доступа к документам в научно-исследовательской организации. Сформулированные требования определили использование различных информационных технологий для их реализации. Общность групп технологий – Web-технологии, протоколы доступа к распределенным базам данных, библиотечные технологии и технологии доставки – позволили выделить в технологии доступа к документам функциональные модули, определяющие информационные процессы технологии. Деление на модули позволяет разделить обслуживание системы доступа к документам между профильными специалистами, осуществлять раздельную модернизацию программных средств и раздельное тестирование модулей для локализации причин сбоев в работе.

В описании информационных ресурсов технологии доступа к документам использован аппарат теории множеств и терминология объектной модели документа. Информационные ресурсы технологии

определены как частный случай объектной модели для единственного класса документов – научных публикаций (или изданий). Показано, что в терминах объектной модели под документом понимаются метаданные издания, описанные в выбранной стандартной схеме метаданных.

В заключение описана реализация технологии доступа к документам для разработки шлюза Z39.50-HTTP Красноярского научного центра СО РАН как единой точки доступа к документам. Построено дерево информационных ресурсов, опубликованных на шлюзе. Шлюз используется для публикации электронных каталогов, библиографических баз данных и тематических коллекций библиотек г.Красноярска. На шлюзе собраны ресурсы, опубликованные с помощью Z39.50. Обоснован выбор программных разработок для реализации отдельных модулей технологии доступа к документам с помощью шлюза КНЦ СО РАН. Показано, что используемое программное обеспечение позволяет логически объединять ресурсы в распределенные базы данных. Для устранения избыточности данных в массивах ресурсов можно произвести их слияние в сводный ресурс с объединением совпадающих описаний экземпляров документов. Показано уменьшение объема сводных ресурсов за счет образования объединенных записей.

На примере различных баз данных проиллюстрирована реализация различных концепций информационных ресурсов, предназначенных для доступа к документам. Доступ к печатным документам, не предназначенным для сканирования, организован как доступ к их аннотациям и информации о наличии документа по месту его хранения. Концепция открытых архивов реализована в архиве научных публикаций организации, где каждый документ описан дважды – в базе данных и как самостоятельная публикация в Web, с использованием схемы Dublin Core. Доступ к лицензируемым документам и документам, доступ к которым ограничен авторским правом, реализован в базе электронных документов. Показано, что дифференциация доступа к документам в такой базе имеет трехступенчатую форму, и может быть построена с учетом всех требований на ограничение доступа.

Результаты работы позволяют рекомендовать разработанную технологию доступа к документам для внедрения в библиотеках и информационных центрах научно-исследовательских организаций, как удовлетворяющую требованиям пользователей и обеспечивающую доступ к разнородным научным публикациям. Технология позволяет реализовать информационные ресурсы, основанные на различных концептуальных подходах, а также использовать её программные средства как инструмент оперативного информирования коллег о результатах научных исследований.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

в журналах, рекомендованных ВАК для представления основных научных результатов диссертации:

1. Ковязина, Е. В. Библиографические базы данных, содержащие ссылки на полный текст документа [Текст] / Е. В. Ковязина // Вычислительные технологии : Специальный выпуск : труды IX рабочего совещания по электронным публикациям «El-Pub2004». - 2005. - Т. 10. - С. 78-81.
2. Ковязина, Е. В. Новые возможности ИРБИС – использование в учебных и практических целях [Текст] / Е. В. Ковязина // Научные и технические библиотеки. - 2005. - № 2. - С. 116-119.

в сборниках и трудах конференций:

3. Ковязина, Е. В. Проблемы создания автоматизированной библиотечно-информационной системы Красноярского научного центра СО РАН [Текст] / Е. В. Ковязина, С. Р. Баженов // Оптимизация информационно-библиографического обслуживания ученых и специалистов : сб. науч. трудов. - Новосибирск, 2000. - С. 28-33.
4. Ковязина, Е. В. Особенности автоматизации библиотек академических институтов [Текст] / Е. В. Ковязина // Библиотеки и ассоциации в меняющемся мире : новые технологии и новые формы сотрудничества : материалы 7-й междунар. конф. «Крым-2000». – М.: ГПНТБ России, 2000. – Т. 1. - С. 204-206.
5. Ковязина, Е. В. Корпоративная библиотечно-информационная система Красноярского научного центра СО РАН [Текст] / Е. В. Ковязина // Научные библиотеки в новом тысячелетии : проблемы взаимодействия ресурсов : материалы регион. науч.-практ. конф. – Новосибирск, 2002. - С. 114-115.
6. Исаев, С. В. Проблемы безопасности библиотечно-информационного сервера [Текст] / С. В. Исаев, Е. В. Ковязина // Библиотеки и ассоциации в меняющемся мире: новые технологии и новые формы сотрудничества : материалы 10-й междунар. конф. «Крым-2003». – М.: ГПНТБ России, 2003. – Т. 1. - С. 258-261.
7. Информационные ресурсы Красноярского научного центра СО РАН [Текст] / В. В. Москвичев [и др.] // Материалы выездного заседания координационного научного совета СО РАН по целевой программе «Информационно-телекоммуникационные ресурсы СО РАН». – Иркутск, 2003. - С. 32-55.
8. Ковязина, Е. В. Электронные ресурсы библиотеки академического института: формирование и использование [Текст] / Е. В. Ковязина // Библиотека и духовная культура нации : материалы регион. науч.-практ. конф. – Новосибирск, 2002. - С. 61-63.
9. Ковязина, Е. В. Формирование электронных ресурсов в библиотеке : проблемы и решения [Текст] / Е. В. Ковязина // Электронные ресурсы региона : проблемы создания и взаимодействия : материалы науч. конф. / Отв. ред. О. Л. Лаврик. - Новосибирск: ГПНТБ СО РАН, 2005. - С. 317-322.
10. Ковязина, Е. В. Использование Z39.50 для публикации электронных каталогов библиотек [Текст] / Е. В. Ковязина // Библиотеки на пути к открытому информационному пространству : материалы научн.-практ. конф. - Красноярск: КрасГАСА, 2003. - С. 24-28.
11. Ковязина, Е. В. Технология заимствования записей - перспективы использования [Текст] / Е. В. Ковязина // Опыт библиотек вузов города Красноярска в области автоматизации и компьютерных технологий : сборник докладов / Отв. ред. Е. Г. Кривоносова, И. С. Колосова. - Красноярск: КрасГУ, 2005. - Вып. 1. - С. 41-45.
12. Ковязина, Е. В. Библиотечно-информационная сеть Красноярского научного центра СО РАН и опыт внедрения ИРБИС в академических библиотеках [Текст] / Е. В. Ковязина // Опыт библиотек вузов города Красноярска в области

- автоматизации и компьютерных технологий : сборник докладов / Отв. ред. Е.Г.Кривоносова, И.С.Колосова. - Красноярск: КрасГУ, 2005. - Вып. 1. - С. 46-50.
13. Ковязина, Е. В. Принципиальная основа, функциональные особенности и информационное наполнение библиотечно-информационной сети КНЦ СО РАН [Текст] / Е. В. Ковязина // Повышение качества непрерывного профессионального образования : материалы всероссийской конф. - Красноярск: КГТУ, 2006. - Ч. 2. - С. 339-343.

другие публикации:

14. Ковязина, Е. В. Автоматизация информационного обслуживания в библиотеке Института вычислительного моделирования СО РАН [Текст] / Е. В. Ковязина. - Красноярск, 1997. - (Препринт ИВМ СО РАН; № 12) - 28 с.

Ковязина Елена Васильевна

ТЕХНОЛОГИЯ ДОСТУПА К ДОКУМЕНТАМ В
НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ
ОРГАНИЗАЦИИ

Автореф. дисс. на соискание учёной степени
кандидата технических наук.

Подписано в печать 20.09.2007.

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.