

Реализация электронной библиотеки на базе типовых модулей распределенной библиотечной системы

Рябев В.С., Соколова Н.В.

Центр "Открытые библиотечные системы"

Санкт-Петербургский государственный технический университет

Санкт-Петербург, Россия

valera@unilib.neva.ru, natalia@unilib.neva.ru

Введение

В результате развития информационных технологий постоянно появляются новые виды информационных объектов, методы и средства их обработки. Функции библиотек не изменились - они по-прежнему заключаются в накоплении, сохранении и предоставлении эффективного доступа к информационным ресурсам. Однако библиотечные технологии претерпевают существенные изменения. Изменился подход к автоматизации библиотек. Если раньше основной целью было создание электронного каталога для своей библиотеки, то сейчас решают проблемы создания распределенных библиотечных систем (БС). Эти системы базируются на глобальной сети Интернет, предоставляют доступ читателя к "виртуальному" пространству, в котором он может найти интересующую его публикацию по сводному каталогу, обратиться к библиотеке-владельцу с заявкой на получение этой книги или послать заказ на электронную доставку фрагмента публикации. Возможно также и получение копии документа из хранилища полнотекстовых документов. Быстрый рост Интернет и развитие технологий мультимедиа в последние годы сделали возможным для широкого круга пользователей использование новых средств работы с информацией. *Реальная библиотека становится "гибридной" - она обеспечивает систематизацию и доступ как к традиционно опубликованным документам так и к электронным коллекциям.*

Россия не занимает лидирующие позиции в области применения ИТ в библиотечном деле. Традиционные для развитых стран технологии: распределенные БС, удаленный доступ, корпоративная каталогизация и др. только начинают внедряться в России. Проекты электронных библиотек (ЭБ) также появляются с некоторой задержкой. Это дает возможность учитывать ошибки и во многом использовать достижения лидеров. Применительно к проектам ЭБ, для нас более актуальна постановка задачи не "как это сделать", а "как это сделать максимально эффективно" и, в первую очередь - с экономической точки зрения. Проекты автоматизации библиотек требуют на начальном этапе существенных материальных затрат, которые не всегда окупаются. Однако, следуя

определенным принципам при проектировании или выборе готового решения, можно обеспечить возможность дальнейшего развития системы с минимальными затратами. Предлагаемым нами вариантом является построение ЭБ на базе распределенной БС. Естественно, не каждая БС способна стать таким фундаментом, поэтому при выборе или при разработке БС надо предусматривать такую возможность ее развития. Современные технологии и существующие библиотечные стандарты позволяют это обеспечить. БС должна иметь возможности для эволюционного развития, т.е. развития, при котором каждое последующее поколение системы наследует существующие компоненты с минимальными изменениями, сохраняя при этом вложенные в БС инвестиции. При этом система должна обладать свойствами функциональной расширяемости и поддержки новых типов данных, пользовательской и платформенной мобильности. *Для этого необходимо, чтобы БС имела открытую архитектуру, т.е. реализовывала открытые спецификации на интерфейсы, протоколы и форматы данных.*

Реализация электронной библиотеки на базе открытой распределенной БС заключается не в создании новой информационной системы, а расширении существующей для новых типов данных. При этом ЭБ наследует основные сервисы БС. Новые сервисы реализуются посредством интеграции в систему дополнительных типовых модулей.

1 Объектная модель данных для электронной библиотеки на базе БС

Построение моделей объектов и процессов их взаимодействия является начальным этапом проектирования сложных систем. Модели позволяют определить необходимый набор компонентов, специфицировать интерфейсы, определить протоколы и форматы данных. Некоторые классы моделей поддаются формальным методам анализа и оптимизации. Как правило, построение моделей систем производится нисходящим методом ("сверху вниз") от наиболее общих прикладных объектов к специфическим компонентам системы, реализующим конкретные функции. Объектная модель предусматривает определение свойств (атрибутов) объектов, методов доступа к ним, а также иерархических и функциональных связей между объектами. Рассмотрим наиболее общую модель взаимодействия пользователей и данных в электронной библиотеке на базе БС.

Первая Всероссийская научная конференция
ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ:
ПЕРСПЕКТИВНЫЕ МЕТОДЫ И ТЕХНОЛОГИИ,
ЭЛЕКТРОННЫЕ КОЛЛЕКЦИИ
19 - 21 октября 1999 г., Санкт-Петербург

1.1 Данные

В ЭБ можно определить следующие основные классы данных:

- Мета-данные
- Данные (электронные документы)

Мета-данные несут информацию о данных, необходимую для их хранения, обработки и использования. В зависимости от назначения мета-данных их также можно разбить на классы. В ЭБ на базе БС можно выделить следующие основные категории мета-данных:

- Ключевые
- Библиографические
- Авторитетные (нормативные)
- Технические
- Библиотечно-учетные
- Условия доступа
- Связи
- Административные

Ключевые метаданные (КМ)

КМ определяют атрибуты объектов, необходимые для идентификации объектов в системе и извне. В общем случае локальные и глобальные идентификаторы объектов могут различаться. Одним из вариантов идентификации объектов является поддержка службы имен [1]. Имена представляют собой глобально уникальные идентификаторы объекта, независимые от текущего местонахождения объекта. Поддержка имен позволяет обеспечить глобальную уникальность, постоянство идентификаторов независимо от перемещения объектов, возможность множественного размещения и доступа к объектам. Одному имени могут соответствовать несколько идентификаторов местоположения (URL) или других имен, определяющих специфические способы доступа к объекту (протоколы). В этом случае КМ могут содержать только один идентификатор DOI (Digital Object Identifier) [2].

Библиографические метаданные (БМ)

БМ содержат библиографические записи (описания). Основное назначение библиографических описаний обеспечить возможность эффективного поиска документа по различным комбинациям полей записи. Некоторые поля описывают содержание документа (предметные рубрики, ключевые слова, аннотация, т.д.). Следует отметить, что в рассматриваемой нами ЭБ на базе БС могут обрабатываться библиографические записи, составленные как для электронных, так и для традиционных документов на твердых носителях (книги, журналы, рукописи и т.д.)

Библиографические записи могут создаваться в различных форматах (семейство MARC, Dublin Core [3]). Может быть несколько вариантов библиографической записи, составленной в различных форматах и/или различными пользователями системы. Например, автор ЭД при размещении его в ЭБ может/должен снабдить его первоначальным описанием в формате Dublin Core, содержащим 15 основных атрибутов документа, таких как заглавие, интеллектуально ответственные лица, тематика и ключевые слова, даты, формат данных и др.

Библиотекарь-каталогизатор, используя данное описание, создает полное библиографическое описание в одном или нескольких MARC форматах. Другой пример: каталогизатор создает RUSMARC описание зарубежного издания книги. При этом он использует библиографическое описание в формате USMARC на английском языке. В каталог могут быть внесены оба описания в различных форматах и с разными языками каталогизации. Преимущества такого подхода для целей поиска и представления документов очевидны.

Авторитетные метаданные (АМ)

АМ (нормативные, пер. с английского "Authority") содержат не описание конкретного документа, а описания некоторых атрибутов описаний, которые могут быть использованы при составлении описаний различных документов. Как правило, это стандартизованные формы представления различных терминов, имен, заглавий. Например, личные имена, наименования организаций, географические названия, унифицированные заглавия, предметные рубрики и др. Использование АМ позволяет сократить временные затраты при составлении описаний, но самое главное - обеспечить единообразное представление часто используемых данных, тем самым значительно повысить качество поиска. Форматы представления авторитетных данных могут быть различными. Это может быть внутрисистемное представление в виде словарей или представление в формате авторитетной MARC записи (USMARC/Authority [4], RUSMARC/Authorities [5], и др.) Очевидно, второй подход имеет преимущество, позволяя обеспечить интероперабельность системы, т.е. возможность общего использования данных и/или взаимного обмена ими с системами других производителей.

Технические метаданные (ТМ)

ТМ задают характеристики документов на электронных носителях необходимые для их обработки программными и техническими средствами. Примерами таких данных являются: формат компьютерного файла (txt, html, bmp, gif, exe и т.д.), размер файла, спецификация местонахождения URL, требования к операционной системе и платформе, формат аудиозаписи и ее продолжительность и т.д. Другими словами, ТМ предоставляют всю информацию необходимую для определения среды и средства доступа к содержанию документа на электронном носителе. Не существует стандартного набора атрибутов, определяющих технические характеристики таких документов. В большинстве случаев для определения соответствующего программного средства для обработки ЭД достаточно расширения файла, в котором хранится документ. Форматы библиографических записей предусматривают определение некоторых характеристик электронных носителей.

Документ на электронном носителе - это документ, доступ к содержанию которого требует специальных технических средств (аудио-, видео- пленки, микрофиши, дискеты, CD-диски и т.д.) Электронный документ, очевидно, имеет электронный носитель. Таким образом, ТМ характеризуют носитель документа. Для ЭД могут содержаться как характеристики самого документа (файла), так и его носителя. Например: dll99.jpg, 300dpi, 200КБ, ГМД 3.5" FAT32. Возможно представление документа на различных носителях и/или разных форматах. Например, один и тот же графический образ может быть представлен несколькими файлами с различным графическим разрешением. Файл с низким разрешением может использоваться для представления образа на WWW-сервере как элемента коллекции образов или ЭД.

По запросу пользователя загружается "полная версия" - файл с высоким разрешением. Другой пример: ЭД, содержащий статью, может быть представлен в нескольких форматах - "html" для просмотра через WWW и архивированный вариант "zip" для эффективного копирования через Internet.

Библиотечно-учетные метаданные (БУМ)

БУМ необходимы для учета документов как единиц комплектования (и/или хранения) библиотечных фондов. Помимо документов на традиционных (бумажных) носителях, такими данными могут обладать документы на электронных носителях (компьютерные дискеты и CD, аудио и видео кассеты, слайды и т.д.). ЭД, размещенный на отдельном носителе (электронная энциклопедия на CD), может обладать такими атрибутами как сигла хранения, инвентарный номер, шифр хранения, количество экземпляров. Для хранения ЭД в "общем" репозитории достаточно наличия ключевых метаданных (например, DOI). Но в общем случае БУМ могут также потребоваться для учета ЭД, размещенных в репозитории ЭД. Например, при приобретении электронной версии журнала через Internet, документ снабжается DOI и размещается в репозитории ЭД. Но при этом факт его приобретения и все необходимые параметры операции должны быть отражены в финансовых документах (учетный документ, дата поступления, источник поступления, цена и др.).

БУМ используются для поддержания традиционного библиотечного цикла (комплектование - каталогизация - хранение - абонемент - списание) и во многом учитывают специфику и технологии этого цикла в конкретной организации. Тем не менее, имеются попытки стандартизации набора таких данных. Примером является формат USMARC/HOLDINGS [6].

Условия доступа (УМ)

Эта группа метаданных необходима для управления правами интеллектуальной собственности на документ (Intellectual Property Rights Management). УМ устанавливает правила использования документа, определяя Кто, какие Операции, и на каких Условиях может выполнять над документом. Атрибут "Кто" может определять категорию пользователей документа (автор документа (владелец авторских прав), читатель библиотеки, внешний пользователь, каталогизатор, администратор репозитория и т.д.). Множество значений атрибута "Операция" зависит от типа носителя документа. Примеры возможных операций с документами на бумажных носителях: заказать по абонементу экземпляр (книги, журнала и т.д.), сканировать N страниц документа. Операции над ЭД: просмотреть, скопировать как файл, получить твердую копию, редактировать содержание, создавать/редактировать метаданные. "Условия" определяют, что должен сделать "Кто", чтобы получить право выполнить "Операцию". Эти атрибуты могут содержать словесное описание процедуры получения прав на использование, стоимость использования (печать или сканирование 1 страницы, копирования 1КБ или всего документа, доступ для чтения в течение 1 часа и т.д.), контактную информацию или ссылку на информационную форму (URL формы регистрации, авторизации, выполнения электронного платежа и т.д.)

Эти метаданные используются, во-первых, для контроля прав доступа к документу и, во-вторых - для поддержки интерфейса между документом и пользователем.

Связи (СМ)

СМ содержат информацию, необходимую для организации связей между документами. Данная группа содержит ключевые атрибуты документа, с которым устанавливается связь (например, DOI) и атрибуты, описывающие характер связи (является частью, содержит часть, ссылка на метаданные, ссылка на содержание документа). Посредством связей можно организовать сложные объекты, состоящие из нескольких, независимо хранимых объектов (коллекция ЭД->ЭД, статья с графическими объектами ЭД->ЭД, ссылки на метаданные: ЭД->БМ, ЭД->БУМ, БМ->БУМ, БМ->АМ и т.д.)

Формат представления СМ и наборы возможных видов связей могут значительно отличаться в различных системах, так как они во многом определяются особенностями реализованной модели данных. Надо отметить, что библиографические форматы семейства MARC имеют внутренние механизмы связи библиографических и авторитетных записей.

Административные метаданные (АДМ)

АДМ содержат служебную информацию, необходимую для управления данными внутри системы. АДМ атрибуты, определяют Кто, Когда и какие Операции выполнял над ЭД или метаданными. Существует стандартный набор административных метаданных A-Core [7]. "Кто" определяется атрибутами: "Name" (имя лица, наименование организации), "Contact" (контактная информация), "Email". "Когда" определяется атрибутом "Date" в формате ISO 8601. Операция задается атрибутом "Activity". Например: created, modified, revised. В случае, если метаданные и их АДМ хранятся в разных объектах, ссылку можно задать в атрибуте "Metadata Location" (например, DOI). Также можно указать лицо, интеллектуально ответственное за создание описываемых метаданных. Кроме того, можно задать время действия АДМ информации в атрибуте dateRange (1999-07-19/1999-09-21).

1.2 Пользователи

Для определения необходимого набора методов доступа к данным рассмотрим основные классы пользователей.

В ЭБ на базе БС можно определить следующие основные классы пользователей данных:

Автор ЭД - лицо, интеллектуально ответственное за создание ЭД,

Читатель библиотеки - лицо, записанное в библиотеку/зарегистрированное в системе,

Внешний пользователь - постороннее лицо, например, получившие доступ к фондам библиотеки посредством Internet

Администратор системы - лицо, осуществляющие техническое обслуживание системы

Библиотечный работник - лицо, осуществляющее обслуживание библиотечных фондов

Библиотечные работники в зависимости от функций, осуществляемых в библиотечном цикле, могут быть разделены на несколько групп. Это разбиение очень сильно зависит от специфики организации цикла в конкретной библиотеке. Для примера выделим наиболее общие группы библиотечных работников:

- Комплектователь (комплектование фондов: заказ, первоначальная обработка новых поступлений, распределение по отделам (сиглам хранения), учет, списание)

- Каталогизатор (ведение каталогов: составление полных библиографических описаний документов, поддержка каталогов АМ)
- Абонемент (поддержка цикла абонемента: учет читателей, обработка заказов, книговыдача)

1.3 Операции

В таблице 1 приведена карта операций, выполняемых различными категориями пользователей над различными классами данных ЭБ. Как отмечалось выше, состав операций зависит от категории данных и типа носителя, а также от специфики реализации модели данных в конкретной организации. Тем не менее, операции над данными могут быть разбиты на две наиболее общие категории:

- приводящие к изменениям данных (W)
- не приводящие к изменениям (R)

Автор при размещении ЭД в ЭБ составляет первоначальное библиографическое описание, описывает технические атрибуты документа и условия использования. Читатель библиотеки при заказе экземпляра документа на твердом носителе фактически изменяет БУМ атрибуты документа (количество доступных экземпляров, количество заказанных и т.д.) Доступ внешнего пользователя к ЭД регламентируется (R?) УМ атрибутами ЭД. Администратор ЭБ имеет полный доступ ко всем метаданным, но ограниченный доступ к содержанию ЭД. Комплекатор, осуществляя размещение/распределение документа в фондах библиотеки, формирует БМ, БУМ, СМ атрибуты документа. Каталогизатор ведет каталоги БМ и АМ. Работник службы абонемента, осуществляя обслуживание читателей, оперирует с БУМ атрибутами документов.

2 Модель контейнера метаданных

Все информационные объекты представляются в БС наборами соответствующих метаданных, а ЭД еще и содержанием. Различные типы объектов могут описываться различными наборами метаданных. Наборы метаданных могут по-разному использоваться разными категориями пользователей. Таким образом, необходимо обеспечить независимый доступ к отдельным наборам метаданных. С другой стороны, для обеспечения целостности объекта метаданные/данные объекта должны быть объединены логически (может быть физически). Другими словами, необходим контейнер, объединяющий различные категории метаданных объекта, с возможностью независимого доступа к ним.

Концепция контейнера метаданных была предложена на Втором семинаре по метаданным (Варвик, Великобритания, апрель 1996г.) и получила название Warwick Framework [8]. На рис. 1 представлена модель контейнера метаданных.

Объект представляется контейнером, в котором содержатся пакеты различных метаданных. Контейнер идентифицируется КМ (например, DOI) Контейнер может быть связан с другими контейнерами посредством атрибутов связи, описанных в СМ пакетах. В общем случае метаданные объекта могут храниться в нескольких контейнерах. Наборы метаданных размещаются в пакетах (package). Пакеты размещаются в контейнере

(container). Возможен независимый доступ к каждому пакету контейнера.

Очевидны следующие достоинства данной модели:

- использование для описания документа нескольких независимых наборов метаданных, каждый из которых определяет только определенную группу свойств (БМ, ТМ, БУМ). Это дает возможность "выделения" стандартных наборов метаданных, понятных всем внешним системам, получающим доступ к данным ЭБ. И в тоже время, создавать свои наборы метаданных, учитывающих специфику конкретной системы.
- разделение ответственности за создание/сопровождение метаданных различных категорий между пользователями (обеспечение защиты данных, эффективная организация цикла обработки данных несколькими категориями пользователей)
- возможность представления разных пакетов в различных/нескольких форматах, что позволяет обеспечить интероперабельность. Различные программные компоненты, как внутренние, так и внешние, имеют возможность работать с теми данными с которыми "умеют" т.е. знают формат представления и значение метаданных и игнорировать все остальные.
- возможность расширения множества используемых метаданных для поддержки новых типов объектов или функциональных возможностей системы

Рассмотрим несколько примеров для наиболее общих категорий объектов "гибридной" библиотеки:

1. ЭД (электронный документ, хранимый в репозитории ЭБ: полный текст статьи, музыкальный файл, программный модуль и т.д.)
2. Традиционный документ (документ представленный на твердом носителе: книга, газета, микрофиша, аудио-кассета и т.п.)
3. ЭД на отдельном носителе (энциклопедия на CD, дискета с примерами, прилагаемая к книге и т.д.)

На рис. 2 приведен пример контейнера метаданных для ЭД, содержащего текст статьи.

В контейнере содержится два пакета БМ для библиографического описания Dublin Core, составленного автором, и описания в формате RUSMARC, составленного каталогизатором. Статья представлена в репозитории ЭБ двумя файлами: в формате pdf размером 12К с идентификатором doi2 и файлом html размером 15К с идентификатором doi3. Условия доступа к документу описаны в объекте doi4. Пакет метаданных АДМ содержит информацию об операциях над пакетами с указанием пользователей и дат. Пакет метаданных связей СМ содержит одно описание связи: объект является частью объекта doi5 (ЭД журнала, в котором опубликована статья).

На рис. 3 приведены примеры контейнеров для книги (переводное издание учебника по HTML) и прилагаемой к ней дискеты с презентацией WWW сайта.

Контейнер книги содержит два библиографических описания: первое в формате USMARC на английском языке, полученное, например, из каталога Library of

Table 1: Карта операций

	КМ	БМ	АМ	ТМ	БУМ	УМ	СМ	АДМ	ЭД
Автор ЭД	R	R/W	R	R/W		R/W			R/W
Читатель	R	R	R	R	R/W	R	R		R
Внешний пользователь	R	R	R	R		R	R		R?
Администратор	R/W	R/W	R/W	R/W	R/W	R/W	R/W	R/W	R
Комплектатор	R	R/W	R	R	R/W	R	R/W		R
Каталогизатор	R	R/W	R/W	R	R	R	R/W		R
Абонемент	R	R	R	R	R/W	R	R		R

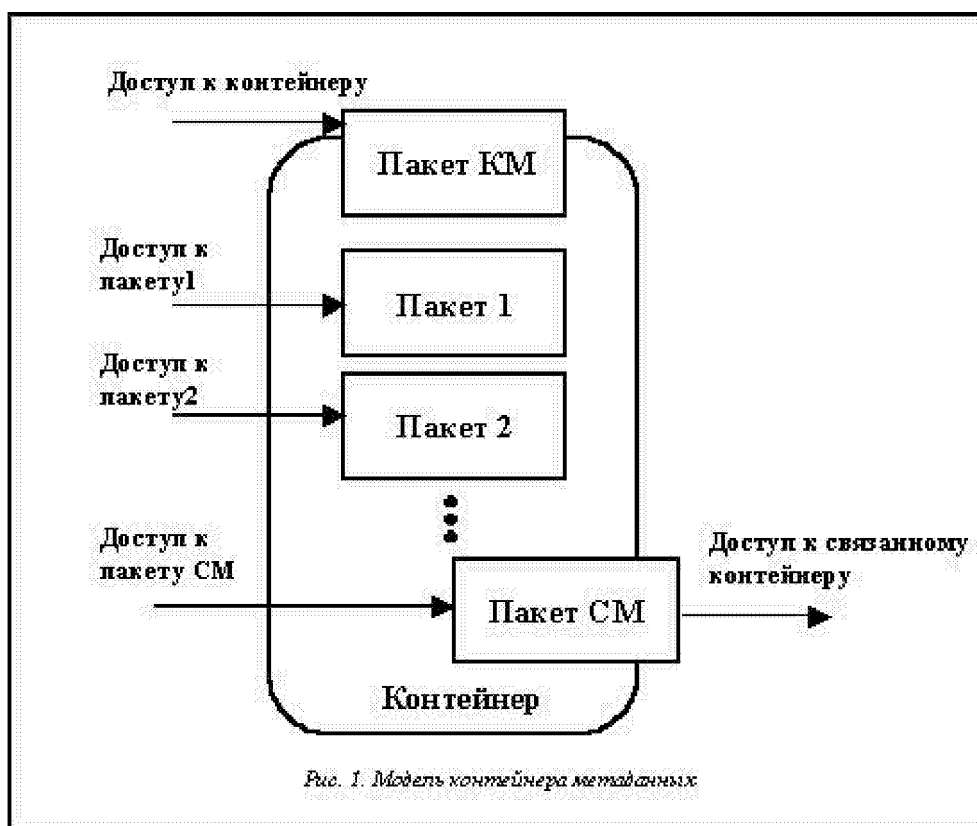


Рис. 1. Модель контейнера метаданных

Congress (USA), и второе в формате RUSMARC, составленное на его основе каталогизатором. Два пакета БУМ содержат информацию о полученных экземплярах книги: пять экземпляров распределено в отдел учебной литературы (ОУЛ) с инвентарными номерами В/У75683 и один в читальный зал (ЧЗ). Пакет СМ содержит ссылку на контейнер прилагаемой дискеты.

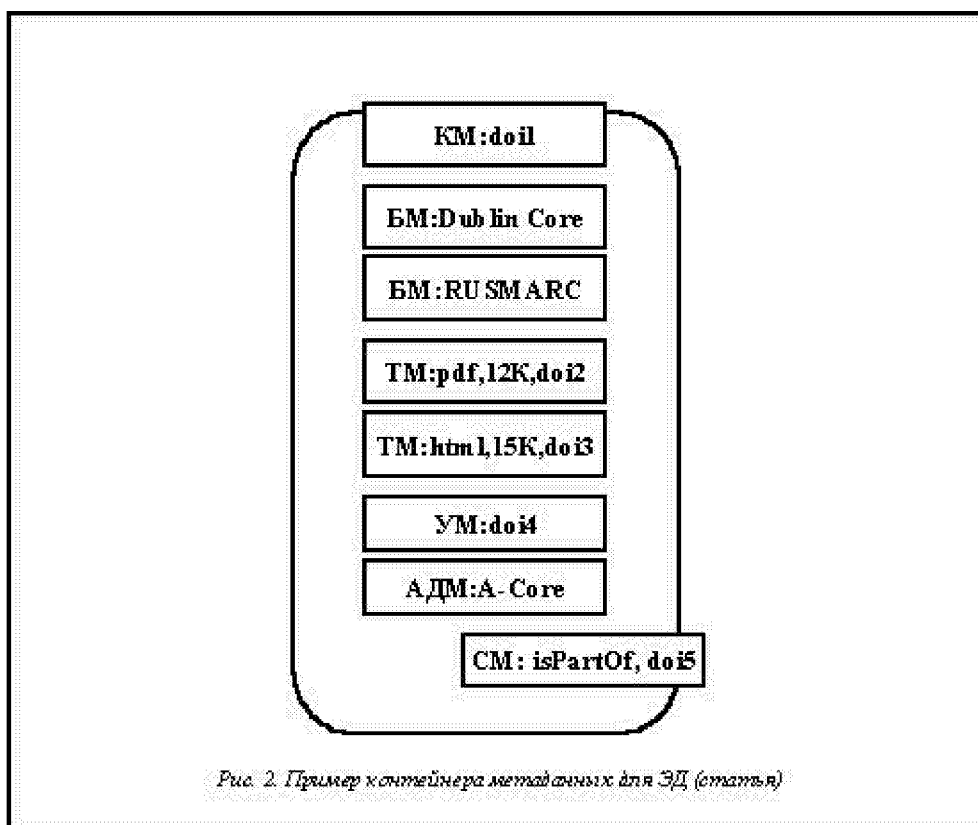
Контейнер прилагаемой дискеты также содержит два пакета БМ: Dublin Core и RUSMARC. Пакеты БУМ содержат информацию о распределении дискет с книгами, но дискеты имеют свои собственные инвентарные номера. Пакет ТМ содержит техническую информацию о ЭД презентации (формат, размер, носитель). Если бы презентация была скопирована в репозиторий ЭД, в контейнер был бы добавлен пакет ТМ с описанием формата, размера файла и DOI. Пакет СМ контейнера дискеты содержит обратную ссылку на контейнер книги.

Таким образом, модель контейнера метаданных обеспечивает унифицированную обработку электрон-

ных и традиционных документов. Различие заключается в наборе пакетов метаданных, которые могут содержаться в контейнере объекта.

3 Реализация модели контейнера метаданных

Для реализации модели контейнера метаданных необходимо представить объект контейнера в виде структуры данных в определенном формате. Формат должен иметь возможность кодирования всех элементов модели: контейнера, пакета, атрибута, значения. Формат контейнера должен быть стандартным, чтобы обеспечить открытый интерфейс к пакетам, которые, как отмечалось выше, могут быть представлены в различных форматах (возможно, специализированных). Возможны несколько вариантов реализации контейнеров. Кратко рассмотрим возможности использования HTML, SGML и MIME форматов и более подробно остановимся на RDF реализации.



3.1 HTML

В данном случае контейнер располагается в блоке заголовка (HEAD) HTML файла. В Приложении 1 приведен фрагмент HTML файла для рассмотренного выше примера ЭД статьи.

Для кодирования атрибутов метаданных используется тег META. Имя атрибута метаданных задается атрибутом NAME тега META, значение - атрибутом CONTENT. С помощью тега LINK задаются ссылки, уточняющие идентификаторы атрибутов, а также ссылки на другие контейнеры.

Данная реализация обладает рядом недостатков:

- избыточность при кодировании атрибутов метаданных (фактически двойное кодирование каждого атрибута)
- нестандартный подход к уточнению идентификаторов атрибутов (SCHEMA)
- невозможность создания более сложных (вложенных) описаний для отражения структуры контейнера, пакетов и сложных атрибутов.

К достоинствам такой реализации можно отнести простоту и возможность прямого встраивания в HTML, что дает возможность инкапсуляции метаданных и содержания ЭД в одном HTML файле (использование метаданных ЭД глобальными поисковыми механизмами, доступ к документу посредством типовой программы просмотра WWW).

3.2 SGML

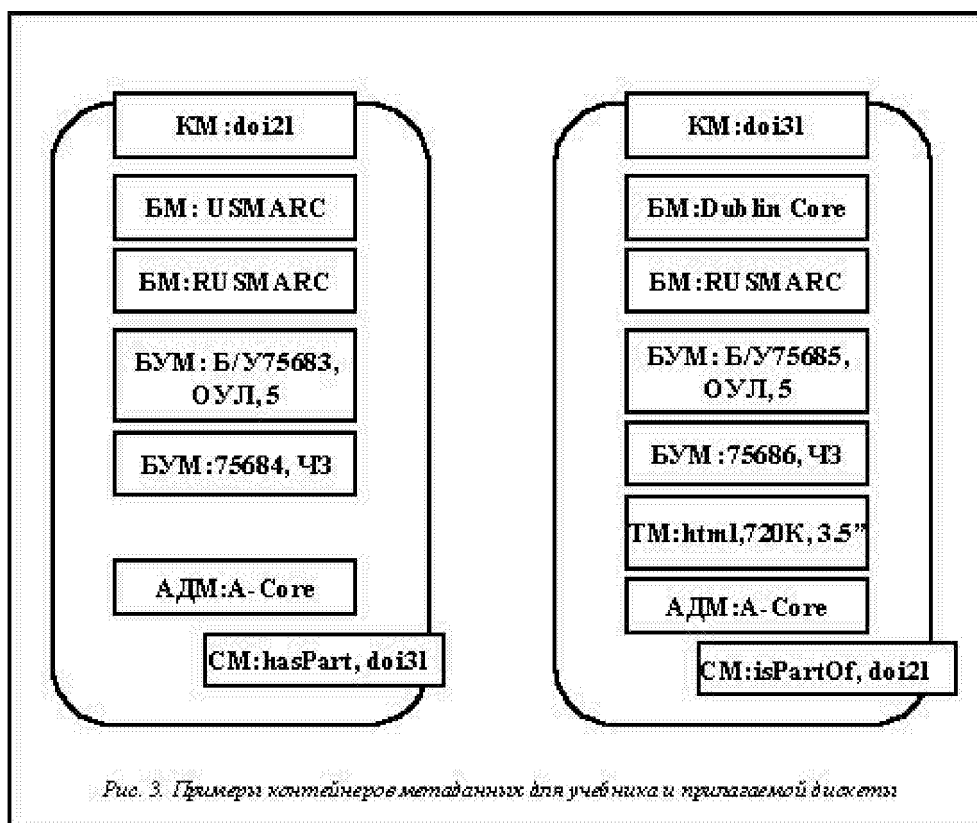
При реализации контейнера на базе SGML [9] можно организовать структуру любой степени сложности за счет определения специфических типов данных (container, package, metadata) Такая реализация требует дополнительного использования файлов DTD (Document Type Definition). Упрощает реализацию возможность использования "готовых" DTD для стандартных форматов (Например, USMARC [10]). Тем не менее, для работы с SGML документами требуются специальные программные средства.

3.3 MIME

MIME (Multipurpose Internet Mail Extentions) [11] - стандартный формат для передачи электронной почты, предусматривающий передачу сообщений в различных нетекстовых форматах. С помощью MIME контейнер может быть представлен как составное сообщение (multipart), части которого (пакеты), могут быть представлены в различных форматах (text/html, text/sgml, application/usmarc и т.д.) MIME является достаточно универсальным средством для кодирования контейнеров метаданных. Однако для работы с MIME форматом также необходимы соответствующие средства.

3.4 RDF

Наиболее универсальным и перспективным форматом для кодирования контейнера метаданных, на наш взгляд, является RDF (Resource Description Framework, W3C



[12]). RDF специфицирует модель и синтаксис для описания различных объектов. Основное назначение RDF - предоставить стандартные возможности для описания ресурсов WWW. Но в модели RDF ресурс рассматривается как абстрактный объект, обладающий свойствами и имеющий идентификатор. Таким образом, контейнер метаданных может быть представлен как набор описаний RDF (rdf:description), каждое из которых содержит пакет метаданных. Описание RDF может иметь сложную структуру (иметь вложенные блоки описаний). Существуют стандартные возможности для группировки, перечисления и выбора элементов описания. Поскольку в качестве базового синтаксиса RDF использует XML [13], имеются возможности указания языка описания и кодировки, что очень полезно, например, при обработке русскоязычных описаний. Кроме того, RDF реализация обладает всеми достоинствами HTML реализации, так как блок RDF может быть напрямую встроен в HTML документ (инкапсуляция с ЭД, стандартные средства обработки, индексирование глобальными поисковыми механизмами). В RDF определены механизмы глобальной идентификации имен атрибутов с использованием пространства имен (xmlns). Таким образом, имеется возможность обработки как стандартных (общеизвестных) наборов атрибутов, например, Dublin Core, так и специализированных, используемых внутри системы. В Приложении 2 приведен пример реализации контейнера метаданных на базе RDF (фрагмент).

4 Программная архитектура ЭБ на базе распределенной БС

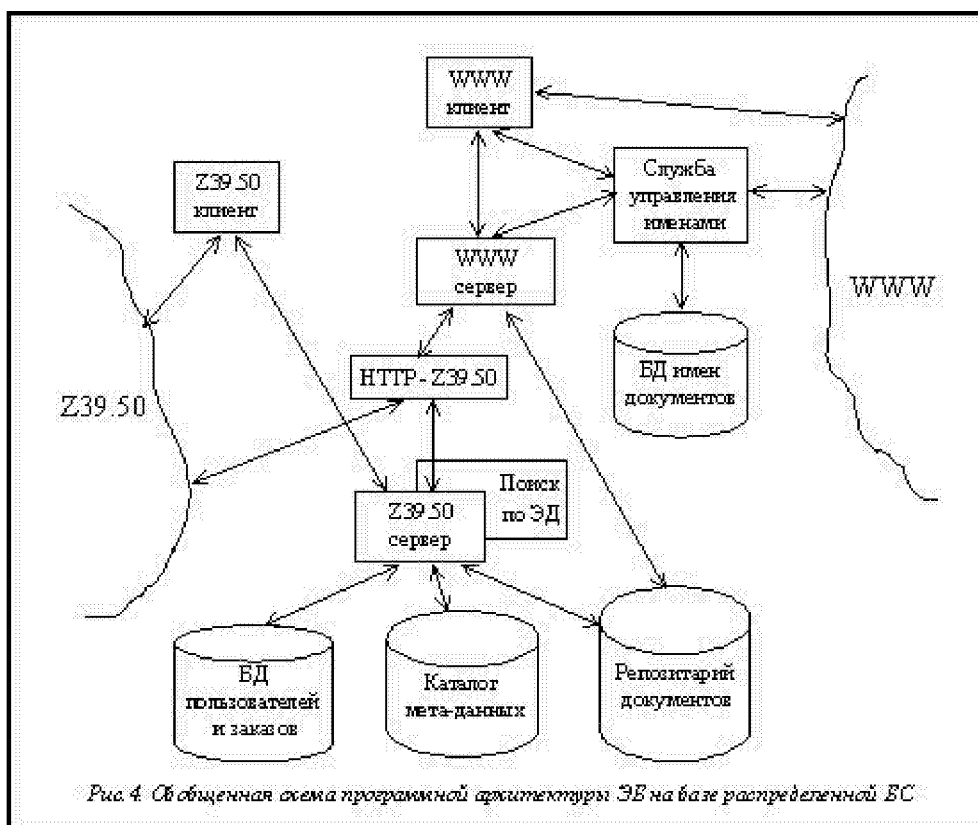
Упрощено ЭБ может быть представлена как совокупность следующих программных подсистем:

- интерфейс пользователя
- поисковая система
- система управления именами документов и контроля доступа
- репозиторий электронных документов (ЭД).

На рис.4 представлена обобщенная схема программной архитектуры ЭБ на базе распределенной БС.

Пользовательский интерфейс реализован на базе WWW и Z39.50 клиентов. Интерфейс работника библиотеки (комплектатор, каталогизатор, служба абонента) обеспечивается семейством профессиональных рабочих мест на базе стандартного протокола Z39.50. Интерфейс читателя/внешнего пользователя реализуется так называемым "тонким" клиентом на базе протокола HTTP, в качестве которого может использоваться типовая программа просмотра WWW.

Поисковая система и все основные функции обработки данных (контроль доступа, каталогизация, заказ, извлечение метаданных и документов) возложены на Z39.50 сервер. Посредством шлюза HTTP-Z39.50 осуществляется взаимодействие WWW и Z39.50 серверов для обслуживания запросов читателей на поиск и извлечение документов. Средствами Z39.50 сервера реализуется



поиск по библиографическим метаданным (БМ). Поисковые возможности системы могут быть расширены за счет включения в систему дополнительных компонентов поиска по содержанию ЭД. Эти компоненты реализуют специальные технологии морфологического, синтаксического и семантического анализа текста.

Новым элементом в архитектуре ЭБ является служба управления именами документов. Как отмечалось выше, имена являются уникальными идентификаторами документов в системе (и в общем случае, должны быть уникальными во всем глобальном информационном пространстве - DOI). Помимо идентификации имени служат для учета авторских прав на ЭД, определения текущего владельца и местонахождения документа в Сети. Имена присваиваются централизованно в рамках системы. Они не зависят от фактического местонахождения документа в Сети. Имена удовлетворяют синтаксису URI, могут содержаться в HTML документах и обрабатываться клиентами WWW. При запросе документа по имени подключается система управления именами, которая преобразует имя документа в URL. Одному имени могут соответствовать несколько URL. Существуют реализации службы в качестве прокси-сервера и расширения типовой программы просмотра WWW.

Репозиторий ЭД, который может содержать информационные объекты различных типов (текст, графика, аудио, видео, программыЖ), реализуется на базе файловой системы. При этом ЭД могут быть доступны для WWW-сервера напрямую. Условия доступа определяются атрибутами УМ. Возможность доступа к ЭД напрямую задается определением в директории службы имен для данного DOI прямой ссылки URL. В случае, если до-

ступ к ЭД ограничен УМ, DOI соответствует другой DOI (контейнер метаданных) с указанием способа доступа через Z39.50 сервер. В этом случае, через HTTP-Z39.50 шлюз, извлекается контейнер метаданных, на основе УМ атрибутов строится HTML форма, через которую на основе метаданных пакета ТМ осуществляется доступ к содержанию ЭД.

Для обеспечения эффективного поиска метаданные ЭД хранятся в отдельном репозитории, доступ к которому осуществляется посредством Z39.50 сервера. Для документов в формате html контейнер метаданных может также копироваться в ЭД для повышения эффективности поиска средствами WWW.

Рассмотренные выше компоненты являются типовыми для современных распределенных БС, реализуют открытые спецификации, могут интегрироваться в систему и при необходимости расширяться в рамках спецификации. На наш взгляд, особого внимания заслуживает Z39.50 клиент. Профессиональному библиотекарю недостаточно функций типового "тонкого" Z39.50 (поиск, извлечение). Для АБИС более целесообразным является создание "толстого" Z39.50 клиента, включающего в себя бизнес-логику всех этапов обработки документов в библиотеке. Он во многом отражает специфику библиотечной технологии в конкретной организации. Однако, этот компонент выполняет вполне определенные функции в рамках системы. Его основная функция заключается в создании метаданных (описаний) информационных объектов, ЭД в том числе. Для обеспечения дальнейшего развития эти приложения должны иметь возможности создания описаний в различных форматах, специфицированных на данный момент времени (семейство MARC,

Dublin Core, IMS, GILS и др.) и адаптации привозникновении новых. Решить данную задачу можно путем обеспечения интерфейса более высокого уровня для создания "описаний форматов описаний". Адаптация к новому формату в этом случае выражается в создании описания нового формата (желательно, чтобы для этого был создан специальный интерфейс пользователя).

Данная модель актуальна при создании на базе распределенных БС электронных библиотек, требующих обработки широкого спектра информационных объектов, представленных в различных форматах.

Заключение

Организация способов эффективного доступа к информационным ресурсам через глобальную сеть Ц область, которая развивается чрезвычайно интенсивно. На наш взгляд, следование современным и перспективным принципам создания сложных систем, обеспечение их расширяемости, масштабируемости, интероперабельности, является единственным приемлемым решением в силу высокой стоимости проектов ЭБ.

Предлагается создание ЭБ на базе современной распределенной БС для традиционных библиотек как подход, позволяющий более эффективно расходовать средства и объединять усилия при решении схожих задач. На основе объектного подхода рассматривается структура документа "гибридной" библиотеки, который может быть элементом хранения, как традиционной библиотеки, так и ЭБ. Предлагается модульная, функционально расширяемая архитектура программной системы, реализующей построение распределенной "гибридной" библиотеки. Предлагаемый подход позволяет сохранять инвестиции в проект распределенной БС и добавлять в нее новые библиотечные сервисы (в данном случае ЭБ), востребованные пользователем.

Библиография

- [1] "Handle System: A persistent Global Name Service Overview and Syntax".- CNRI, July 16, 1998.
- [2] *The Digital Object Identifier System*, <http://www.doi.org/>
- [3] *Dublin Core Metadata Initiative*, <http://purl.org/DC/>
- [4] *The USMARC Format for Authority Data*, <http://lcweb.loc.gov/marc/authority/ecadhome.html>
- [5] *Российский коммуникативный формат представления авторитетных/нормативных записей*, <http://www.nlr.ru/rba/auth/content.htm>
- [6] *The USMARC Format for Holdings Data*, <http://lcweb.loc.gov/marc/holdings/echdhome.html>
- [7] "The A-Core: Metadata about Content Metadata", Internet-Draft//R. Iannella, DSTC, D. Campbell, NLA, 30 June 1999, <http://metadata.net/admin/draft-iannella-admin-01.txt>
- [8] "The Warwick Framework: A Container Architecture for Diverse Sets of Metadata"//Carl Lagoze, Digital Library Research Group, Cornell University.- D-Lib Magazine, July/August 1996, <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>
- [9] *ISO (International Organization for Standardization). ISO 8879:1986(E). Information processing – Text and Office Systems – Standard Generalized Markup Language (SGML)*. First edition – 1986-10-15. [Geneva]: International Organization for Standardization, 1986.
- [10] *MARC DTDs (Document Type Definitions)*, <http://lcweb.loc.gov/marc/marcsgml.html>
- [11] *MIME (Multipurpose Internet Mail Extensions)*.- RFC-1341, RFC-1521, RFC-1522
- [12] "Resource Description Framework (RDF): Model and Syntax Specification", W3C Recommendation, 22 February 1999, <http://www.w3.org/TR/REC-rdf-syntax/>
- [13] "Extensible Markup Language (XML) 1.0", W3C Recommendation, 10-February-1998, <http://www.w3.org/TR/REC-xml>

Приложение 1

```
<HTML> <HEAD>
<TITLE> Пример реализации контейнера метаданных на базе HTML </TITLE>
<LINK REL="SCHEMA.dc" HREF="http://purl.org/DC/about/element\_set.htm" >
<LINK REL="SCHEMA.olsc" HREF="http://www.unilib.neva.ru" >
<LINK REL="SCHEMA.ims" HREF="http://www.imsproject.org/metadata/" >
<META NAME="olsc.doi" CONTENT="doi1">
<META NAME="dc.Title" CONTENT=" Реализация электронной библиотеки на базе типовых модулей распределенной БС ">
<META NAME="dc.Creator" CONTENT="Соколова Н.В.">
<META NAME="dc.Creator" CONTENT="Рябев В.С.">
<META NAME="olsc.rusmarc" CONTENT="00621nam2200193345000 : MARC-запись :">
<META NAME="ims.technical.format" CONTENT="pdf">
<META NAME="ims.technical.size" CONTENT="12K">
<META NAME="ims.technical.location" CONTENT="doi2">
<META NAME="ims.technical.format" CONTENT="html">
<META NAME="ims.technical.size" CONTENT="15K">
<META NAME="ims.technical.location" CONTENT="doi3">

... атрибуты УМ, АДМ и СМ ...

</HEAD> <BODY>
Текст статьи
</BODY> </HTML>
```

Приложение 2

```
<?xml version='1.0' encoding="Win1251" ?>
<RDF xmlns="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/DC/elements/ "
xmlns:ac="http://metadata.net/ac/2.0/"
xmlns:p="http://www.unilib.neva.ru/"
xmlns:ims="http://www.imsproject.org/metadata/">
<Alt>
<li><Description about= "doi1" xml:lang="rus" bagID= "p1">
  <dc>Title> Реализация электронной библиотеки на базе типовых модулей распределенной БС " </dc>Title>
  <dc:Creator>
  <Bag>
  <li> Соколова Н.В. </li>
  <li> Рябев В.С. </li>
  </Bag>
  </dc:Creator>
</Description></li>
<li><Description about= "doi1" bagID="p2">
  <value> 00621nam2200193345000 : MARC-запись :</value>
</Description></li>
</Alt>
<Bag>
  <li><Description about= "doi1" bagID= "p3">
  <ims:Technical>
  <ims:Format> pdf </ims:Format>
  <ims:Size> 12K </ims:Size>
  <ims:Location> doi2 </ims:Location>
  </ims:Technical>
  </Description></li>
  <li><Description about= "doi1" bagID="p4">
  <ims:Technical>
  <ims:Format> html </ims:Format>
  <ims:Size> 15K </ims:Size>
  <ims:Location> doi3 </ims:Location>
  </ims:Technical>
  </Description></li>
</Bag>
:
<Description aboutEach= "\#p2">
<p>Type> BibRecord </p>Type>
```

```
<p:Schema> Rusmarc </p:Schema>
<p:Format> ISO2709 </p:Format>
<ac:Name> CatLibrarian </ac:Name>
<ac:Email> CatLib@unilib.neva.ru </ac:Email>
<ac:Activity> created </ac:Activity>
<ac>Date> 1999-08-25 </ac>Date>
</Description>
:
</RDF>
```