

Интегрированная Система Информационных Ресурсов (ИСИР) РАН

- подход к созданию интегрированных цифровых библиотек

Агошков С.В., Бездушный А.Н., Галочкин М.П.,
Кулагин М.В., Меденников А.М., Серебряков В.А.
e-mail: smsall@ccas.ru

1 Введение

Информационно-поисковые системы играют центральную роль в поиске информации в цифровых библиотеках и Интернет. Однако, поисковые системы Интернет не позволяют осуществить эффективный поиск требуемой информации. Это связано и с возрастанием объема информации, и со способами выбора того, что следует индексировать, как обеспечить единое пространство, с проблемами определения, в контексте каких запросов следует выдавать ту или иную информацию.

Основное различие между специализированными поисковыми системами, обеспечивающими доступ к предметно-ориентированным ресурсам, и популярными Интернет системами состоит в качестве результатов, получаемых пользователями. Специализированные системы дают достаточно точные результаты, предоставляют его в удобной для чтения форме, тогда как результаты поиска в Интернет системах, представляются как 'сырые' данные, содержат массу лишних документов. Это определяется природой процесса каталогизации и индексации информации. Специализированные системы каталогизируют отобранные подготовленные информационные ресурсы, индексируют их, используя специально подготовленные описания ресурсов (метаданные), а Интернет системы осуществляют автоматическую индексацию Web-ресурсов на основе некоторых слов документов.

Попытки централизованно создать каталоги ресурсов Интернет не дают приемлемого решения. Они не полны, не точны, поскольку Интернет постоянно меняется. Создаются одни ресурсы, удаляются другие, меняется содержание третьих. Любой каталог без надежного механизма поддержки актуализации данных быстро устаревает. Попытки отказаться от централизованной каталогизации, дав создателям Интернет ресурсов средства самокаталогизации (внесения сведений о ресурсе в некоторый каталог), тоже трудно назвать успешными, поскольку невозможно зарегистрироваться во всех каталогах, имеющих к тому же собственные форматы описания каталогизируемых ресурсов.

Поскольку ни одна из поисковых машин не индексирует все содержимое Интернет, пользователи должны

обращаться к разным поисковым системам, которые имеют разные возможности, разные поисковые интерфейсы. Желательно автоматизировать этот процесс, но не существует общепринятых стандартов на передачу поисковых запросов одновременно нескольким поисковым машинам и последующую достаточно разумную аккумуляцию их ответов.

Как правило, поисковые системы индексируют текстовые документы. Считается, что интеллектуальное содержание, прежде всего, представлено текстом, хотя информационные ресурсы могут иметь сложное строение, включать разнообразные бинарные данные. Даже текстовые документы желательно различать по их специфике: одни являются статьями, другие - отчетами, третьи описывают реальные или виртуальные объекты (программы, персон и т.п.). Одна информация должна интегрироваться с другой связанной с ней информацией. Для обеспечения интеграции необходимо использовать средства идентификации ресурсов.

В связи с вышеизложенным, приступая к реализации информационной системы РАН, мы выделили следующие ключевые моменты и направления реализации распределенной системы с открытой компонентной архитектурой - интеграция разнотипных ресурсов, их идентификации, использование метаданных, применение открытых стандартов взаимодействия систем, поиска, обмена и представления данных.

1.1 Цели проекта

Основной целью проекта ИСИР РАН является разработка концептуальной основы и инфраструктуры для интеграции разнородных информационных и вычислительных ресурсов РАН в единое информационное пространство. Этот базис должен обеспечить объединение в единое пространство всевозможных цифровых библиотек, информационных и вычислительных систем, использующих как собственные принципы организации, так и технологию открытой архитектуры системы проекта ИСИР или непосредственно ее релизы.

Гибкая организация информации, ее интегрированное представление, открытая архитектура системы являются ключевыми моментами в проектировании системы ИСИР.

К системе предъявлялись следующие требования:

- Логическая группировка данных - система должна позволять обрабатывать все запросы на логических группах баз данных, полностью скрывая тем самым физическое расположение последних.

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

- Абстрактная модель данных - информационная система строится на основе абстрактной схемы данных, на которую должны быть отображены конкретные базы данных, что позволяет объединять данные из разнородных систем в одной логической группе.
- Абстрактная система запросов - система должна оперировать не конкретным синтаксисом запросов, а его логической сутью на основе абстрактных ресурсов и их атрибутов.
- Метаинформация - система должна владеть полной информацией о себе и обо всех своих ресурсах.
- Разграничение доступа - система должна быть способна предоставлять различные уровни доступа к информации для различных пользователей.
- Учет и контроль - система должна уметь собирать статистику по запросам пользователей и вести их бюджеты.
- Открытость - система должна легко расширяться и быть основана на открытых стандартах и протоколах.
- Работа с распределенными данными - информационная система должна допускать возможность работы с данными, расположенными на разных физических серверах, различных аппаратно-программных платформах.
- Связь с другими системами - возможность системы интегрировать свои ресурсы с ресурсами других информационных систем и взаимодействовать с другими при осуществлении поиска информации.
- Легкость в общении - для пользователей система должна предоставлять простые интерфейсы доступа к информации. Необходимым элементом системы сегодня является также наличие WEB-доступа.

1.2 ИСИР как цифровая библиотека

Мы рассматриваем систему ИСИР РАН как цифровую библиотеку, поскольку, как и цифровые библиотеки, система предназначена для хранения и обеспечения взаимодействия с большими распределенными массивами разнообразных ресурсов. Как и цифровые библиотеки, система призвана обеспечить следующие потребности пользователей:

1. Хранение всевозможной информации в электронном виде (текст, сканированный текст, графические изображения, звук, видео, программы). Обеспечивать механизмы загрузки разного рода информации.
2. Распределение информации по сети.
3. Поиск по распределенной информации.
4. Выдачу (в обобщенном смысле) информации.
5. Соблюдение прав доступа и защиту информации.
6. Механизмы загрузки, поиска и выдачи информации должны поддерживать взаимодействие с разнородными электронными библиотеками (как внутри страны, так и за рубежом).

Однако, мы используем более широкое толкование понятия цифровой библиотеки и его предмета и свойств. Мы считаем, что *содержание* библиотеки составляют не просто *информация* (цифровые объекты), а **ресурсы**, которые могут быть как информационными и программными, так и вычислительными ресурсами. Ресурсы относятся к разнообразным **типам**, а не являются только документами (в широком смысле этого слова), пусть даже мультимедийными и со сложным описанием и строением. Эти разнородные ресурсы должны составлять **интегрированное** пространство библиотеки. Они тесно взаимосвязаны - одни характеризуют другие, вторые воздействуют на первые и т.п. Мы исходим из того, что *функциональность* библиотеки при работе с ресурсом не исчерпывается предоставлением *доступа* к информации (ее извлечением и представлением). Пользователям может быть предоставлена возможность **взаимодействовать** с обнаруженными ресурсами. Точнее, взаимодействовать с объектами, обнаруженных ресурсов в соответствии с их открытыми интерфейсами. Это обстоятельство может быть использовано компонентами библиотеки, что позволит ей существенно расширять набор предоставляемых услуг.

Концептуальная модель системы не нарушает основополагающие принципы цифровых библиотек, изложенные в работах [1] и [2], в которых были определены основные понятия цифровой библиотеки:

- '**цифровой объект**' (digital object) - структура данных, основными компонентами которой являются *данные* и *ключевые метаданные*, включающие уникальный идентификатор;
- '**уникальный идентификатор**' (handle, URN), идентифицирующий цифровой объект в сетевой распределенной среде;
- '**хранилище**' (repository), сохраняющее цифровые объекты.

Мы стремились с одной стороны расширить, а с другой конкретизировать эти понятия так, чтобы они представляли не просто распределенные однородные коллекции цифровых объектов, а описывали распределенное *интегрированное* пространство *электронных* ресурсов, которые не только вместе хранятся, но характеризуют друг друга, и которые являются не просто последовательностями бит, но и функциональными компонентами, способными расширить набор стандартных служб библиотеки, предоставить пользователям искомые *программные* сервисы или *вычислительные* ресурсы [8].

2 Основные понятия цифровой библиотеки

В цифровой библиотеке логическая единица хранения называется **ресурсом** или объектом. Ресурс имеет две части - **метаданные** (*структурированные сведения о ресурсе*) и, возможно, **данные** (*содержание ресурса*). Метаданные - это сведения о ресурсе, на основе которых осуществляется управление ресурсом, его поиск и взаимодействие с ним. К метаданным относится и **уникальный идентификатор** ресурса, обеспечивающий идентификацию ресурса в соответствующей среде - локальной или распределенной сетевой среде. Интерпретация содержания ресурса осуществляется на основе его типа или программного интерфейса, описываемых в метаданных. Ресурсы библиотеки *не однородны*. Это проявляется

в различии их свойств, предоставлении услуг. Каждый ресурс относится к некоторому типу, который называется **типом ресурса** и определяет:

- виды сведений о ресурсах этого типа, их услуг (состав и структуру метаданных);
- механизмы поиска ресурса и представления информации о нем;
- возможные взаимодействия (связи) с другими ресурсами;
- способ внутренней организации ресурса;
- методы взаимодействия ресурса с пользователем.

Ресурсы тесно **взаимосвязаны** друг с другом, что обусловлено как взаимными характеристиками (*персона* является автором *публикации*, которая подготовлена в рамках *проекта*, спонсированного *организацией*), так и взаимными действиями (*программа* транслируется *компилятором*, работает под управление *ОС* и обращается к *процессам*). Связи, как и ресурсы не однородны и типизированы. Выбор набора типов ресурсов зависит от ориентации цифровой библиотеки, ее предметной области. Цифровая библиотека должна поддерживать следующую функциональность:

- предоставление сведений о ресурсах,
 - поиск ресурсов на основе метаданных, в частности полученных в результате обработки полнотекстовых документов,
 - навигация в пространстве связанных ресурсов,
 - просмотр свойств ресурсов,
- управление ресурсами,
 - загрузка метаданных и содержания ресурсов из разнообразных форматов,
 - выгрузка метаданных и содержания ресурсов в разнообразные форматы,
 - автоматизированная подготовка метаданных,
 - публикация, удаление ресурсов, редактирование их свойств и, возможно, самих ресурсов,
 - индексирование метаданных,
 - поддержка средств идентификации ресурсов,
 - поддержка актуальности ресурсов,
 - управление доступом к ресурсам и аудит доступа,
- взаимодействие с ресурсами,
 - извлечение ресурсов, предоставление содержания ресурса,
 - поддержка и обеспечение функциональности ресурсов на основе открытых интерфейсов,
 - обработка (возможно, аналитическая) информационного наполнения.
- поддержка распределенности ресурсов,
 - аккумуляция и кэширование метаданных,
 - перенаправление, широковещательная рассылка поисковых запросов,

- интеграция и идентификация сторонних ресурсов,
- актуализация информации о сторонних ресурсах,
- мета-поиск (поиск в среде цифровых библиотек и поисковых систем иной структуры),
 - преобразование поисковых запросов,
 - интегрирование результатов ответов на поисковые запросы,

Основные из этих функций определяются ниже и иллюстрируются на рис. 1:

- *Публикация* - процесс предоставления ресурса некоторым пользователем системы, в результате которого другие пользователи могут найти этот ресурс и обратиться к нему.
- *Резюмирование* - процесс извлечения представительного множества характеристических сведений (метаданных) из ресурса или его окружения.
- *Индексирование* - процесс перевода информации, полученной после резюмирования, в форму, обеспечивающую эффективный распределенный поиск.
- *Поиск* - процесс обработки 'индекса', имеющий целью сформировать ответ на запрос пользователя по распределенным метаданным.
- *Просмотр* - контролируемая пользователем деятельность, направленная на обследование распределенного информационного пространства, формируемого системой на основе ресурсов и их связей.
- *Извлечение* - деятельность пользователей, вызванная желанием получить (извлечь из распределенного информационного пространства) обнаруженную информацию.
- *Взаимодействие* - процесс, ведомый пользователем и находящийся в рамках, определяемых ресурсом, в течение которого ресурс предоставляет пользователю информационный или функциональный сервис.
- *Распределение* - деятельность системы, имеющая целью обеспечение работы с данными, расположенными на разных физических серверах, различных аппаратно-программных платформах.
- *Преобразование* - деятельность системы, направленная на перевод метаданных других систем, хранящихся в различных внутренних форматах, во внутренний формат системы для обеспечения интеграции информации и эффективного поиска.
- *Сотрудничество* - деятельность системы, ориентированная на совместную деятельность с другими системами по обслуживанию поисковых запросов пользователей, в основном заключающаяся в преобразовании поисковых запросов для этих систем и интегрировании результатов их ответов.
- *Контроль* - деятельность системы по предоставлению различных уровней привилегий доступа к информации, ее защите, по ведению бюджетов пользователей.

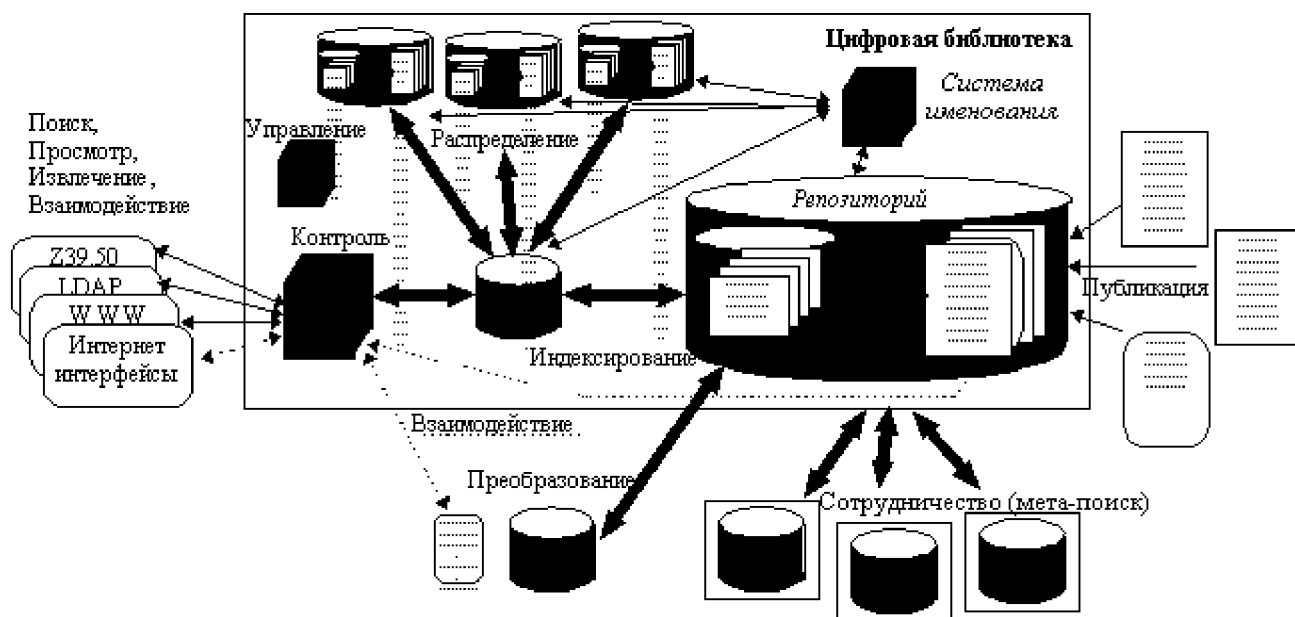


Рис. 1:

- *Управление* - деятельность по сопровождению системы, ее ресурсов, соблюдению актуальности, целостности и сохранности информации в распределенном информационном пространстве.

2.1 Ресурсы

Ресурс - это некоторая условная единица информации, которая может иметь как достаточно тривиальную структуру, так и довольно сложное внутреннее строение. Ресурсом может быть все, что обладает индивидуальностью. Не всякий ресурс имеет электронную основу (однако, может иметь электронный образ), следовательно, содержание не всякого ресурса можно получить по сети. Выбор того, что является ресурсом, а что его свойством, зависит от предметной области, от целей, преследуемых приложениями, обслуживающими ресурсы, и ряда других причин. Например, при рассмотрении организационной структуры в случае межкорпоративной системы подразделения организаций могут выступать в качестве свойств ресурса 'организация', а в случае внутрикорпоративной системы - как самостоятельные ресурсы.

2.2 Метаданные

Метаданные - это сведения о данных, свойства данных. Метаданные могут иметь разнообразное назначение, в том числе они могут:

- определять диапазон возможностей поиска ресурса (например, заглавие, автор, ключевые слова публикации, ФИО и должность сотрудника, название и тип организации и т.д.)
- определять возможности навигации в пространстве соответствующих ресурсов (например, по темам публикаций, по должностям сотрудников, по иерархии административного подчинения организаций и т.д.)
- характеризовать правила работы с ресурсами данного типа (например, указывать на то, что книга состоит из глав, делится на страницы и т.д.)
- предоставлять разнообразную административную информацию о ресурсе (например, время создания, права доступа, кто изменил и т.д.)

Метаданные характеризуют ресурс посредством набора пар (**атрибут, значение**), где атрибут именуется одно из ключевых свойств ресурса, а значение описывает это свойство. Значения атрибутов могут иметь сложную организацию, соответствующую структуре свойства - его подразделению на части, взаимному влиянию частей.

В зависимости от ориентации, уровня абстракции цифровой библиотеки одни и те же информационные объекты могут выступать как в качестве свойства некоторого ресурса, так и в виде самостоятельного ресурса, связанного с первым (например, статья - издание, подразделение - организация, модуль - программа). От предметной области и уровня детализации цифровой библиотеки зависит также выбор набора свойств, представляющих ресурсы.

В соответствии с использованием метаданных службами библиотеки элементы метаданных образуют пересекающиеся группы, которые можно классифицировать следующим образом:

- Метаданные для **поиска** представляют те элементы метаданных, которые наиболее часто будут использоваться пользователями при формировании поисковых запросов или в процессе навигации в информационном пространстве библиотеки. Например, название, ключевые слова описания, дата.
- Метаданные для **просмотра** включают в себя метаданные для поиска и дополнительные элементы, которые следует предоставлять пользователям в результатах поисковых запросов или в процессе навигации в информационном пространстве библиотеки,

чтобы пользователи смогли определить, является ли рассматриваемый ресурс искомым или потенциально таковым. Разным видам просмотра сведений соответствуют разные наборы элементов метаданных:

- **Краткие сведения** выводятся в результатах поисковых запросов или в процессе структурной или тематической навигации, поскольку одновременно предоставляются сведения о достаточно большом числе ресурсов.
 - **Стандартные сведения** включают представительный набор характеризующих ресурс свойств, позволяющих пользователям детально познакомиться с ресурсом.
 - **Развернутое описание** ресурса содержит полное перечисление всех элементов метаданных, необходимое владельцу ресурса или администратору системы для контроля и анализа описательных и функциональных характеристик ресурса. В эту форму могут включаться и разнообразные вычисляемые элементы, статистического, аналитического характера.
- **Метаданные для редактирования** описывают, какие из сведений о ресурсе (включая его содержание) можно изменять, а какие можно только уточнять (расширять). Они указывают можно ли, когда, при каком условии удалить ресурс; что должно происходить при его удалении, какие действия может выполнять пользователь, выступающий в определенной роли.
 - **Метаданные управления доступом** определяют права на использование, категории пользователей, которые могут выполнять поисковые запросы, находить ресурсы, просматривать или редактировать их в том или ином объеме. Управление доступом, например, может быть осуществлено с помощью типизации доступа к ресурсам - публичный ресурс, условно доступный ресурс, ресурс с уведомлением доступа, ресурс с регистрацией доступа, ресурс с платным доступом и т.п.
 - **Метаданные идентификации** ресурса характеризуют уникальность ресурса (а не одной из его форм представления) в пределах коллекции библиотеки или информационного пространства, в котором она находится.
 - **Метаданные актуализации** ресурса описывают период актуальности сведений и данных ресурса, определяют, что следует делать после истечения срока, как восстановить актуальность ресурса.
 - **Метаданные для управления процессом взаимодействия** с 'данными' ресурса, его **функциональностью** отвечают за функциональность самого ресурса. В простом случае это извлечение и отправка пользователю данных ресурса, например, PDF-файла. В более сложном случае за этим скрывается преобразование 'данных' ресурса (набора бит) в функциональный объект (объект в смысле ООП) - компонент библиотеки, программные интерфейсы которого описываются этими метаданными.

Можно видеть, что по большому счету имеется два вида элементов метаданных ресурсов:

- **описывающие ресурс** - представляют внутренне присущие ресурсу свойства,
- **управляющие ресурсом** - обеспечивают управление и взаимодействие с ресурсом. Некоторые из элементов метаданных этого вида могут характеризовать поведение всех ресурсов одного и того же типа, а не каждого из них в отдельности.

2.3 Отношения между ресурсами

Существует множество форматов метаданных для описания разнообразных видов информации - публикаций, программ, химических формул, географических данных и т.п. Наверное, наибольшее число форматов метаданных, называемых библиографическими описаниями, имеется для публикаций. Это RFC 1807, BibTex, IAFA, DC, GILS, всевозможные вариации формата MARC (ISO2709, USMARC, UNIMARC, RUSMARC) и т.д [13, 9]. Каждый из них в определенной мере, контексте и для определенной цели по-своему удобен и хорош.

В Интернет среде и в среде электронных библиотек все возрастающей популярностью пользуется формат DC (Dublin Core) [3, 7], предназначенный для описания разнообразных электронных объектов (публикаций, программ, видео и т.п.). Анализируя этот простой формат, можно видеть, что его элементы имеют разную природу - одни описывают внутренние свойства рассматриваемого ресурса, а вторые характеризуют другие информационные ресурсы, точнее, связи с ними. На рис. 2. в графическом виде показано разбиение элементов формата DC на атрибуты публикации и ее связи с другими информационными ресурсами.

Можно сделать вывод о необходимости хранения и работы с метаданными не как с совокупностью сведений об одном типе информационных ресурсов, а как с совокупностью сведений о ряде взаимосвязанных типов информационных ресурсов. Соответствующая этому обработка метаданных (рис. 3) позволяет на том же самом наборе метаданных обеспечить следующие возможности:

- выполнять более *точные* и более *сложные* поисковые операции,
 - в качестве результата поиска получать сведения не только о публикациях, но о *связанных с ними ресурсах* и много другой информации, например, *аналитического характера*.

Включение в это множество типов ресурсов, не указываемых в библиографических описаниях, но часто связанных с научными публикациями, таких, как проекты, конкурсы, гранты, конференции, спонсоры, организации, позволяет, например, получить схему связи типов ресурсов, изображенную на рис. 4.

Такая система типов и связей ресурсов существенно расширяет возможности информационной системы по предоставлению разнообразных сервисов. Например, можно осуществить следующие виды запросов:

- по автору публикации узнать, в каких проектах он принимает участие, и посмотреть материалы и публикации этих проектов;
- узнать о конференциях, на которые автор представлял свои публикации;

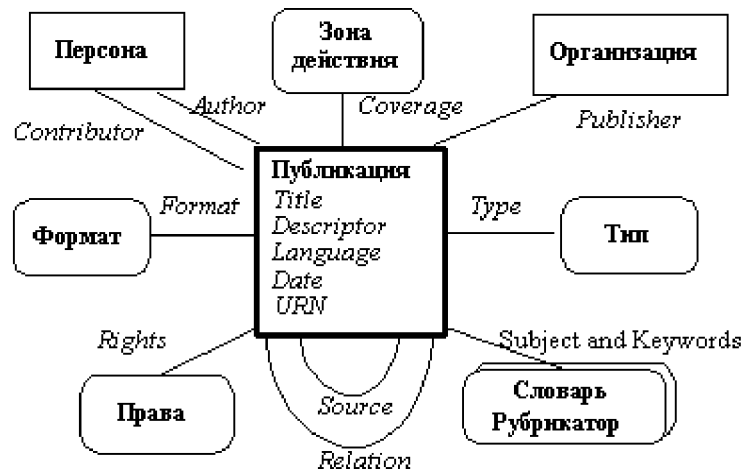


Рис. 2:

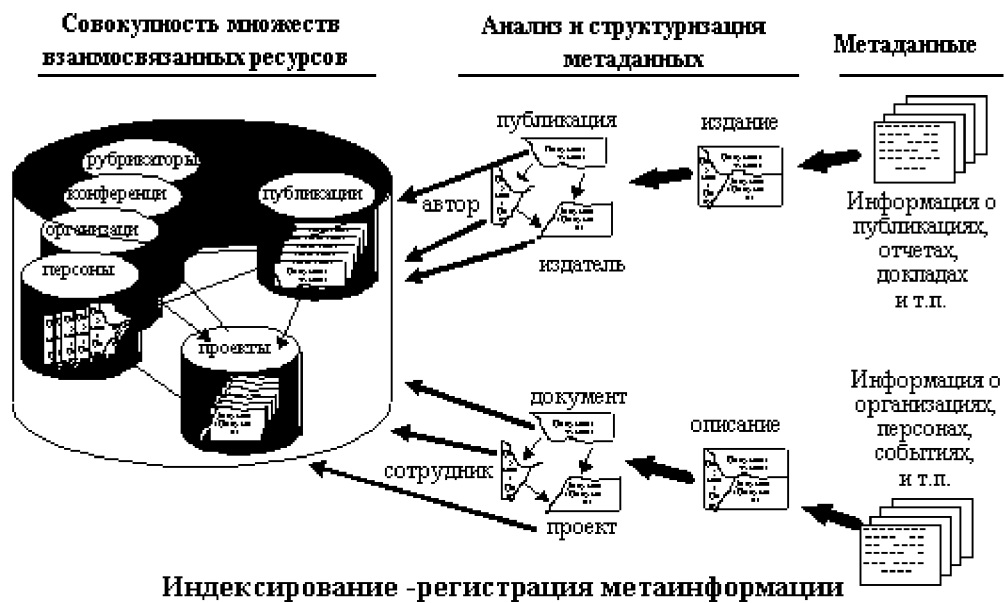


Рис. 3:

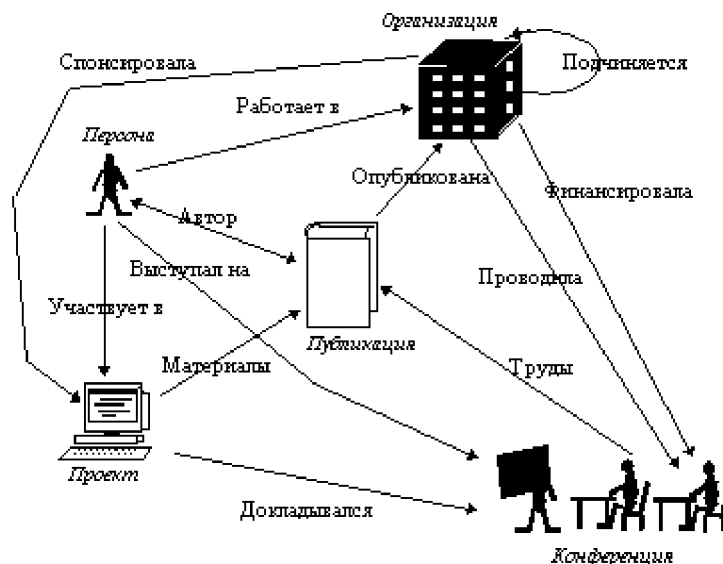


Рис. 4:

- найти публикации коллег автора по работе;
- познакомиться с материалами тех секций конференций, на которых коллеги автора по проекту докладывали о своей работе.

Видим, что ресурсы цифровой библиотеки сильно связаны, поскольку при описании их свойств имеет место взаимная характеристика одних другими, а при предоставлении свойственных ресурсам услуг они используют друг друга, воздействуют друг на друга и т.п. Традиционно этот вид свойств ресурсов представляется атрибутами метаданных с помощью примерного описания ресурсов, связанных с рассматриваемым ресурсом. Это не желательно, но допустимо при обмене метаданными между разнородными системами. В современных цифровых библиотеках, оперирующих с огромными объемами информации, обслуживающих массы пользователей, использование приближенных описаний неприемлемо.

Необходимо выделить из метаданных описание взаимосвязи (**отношений**) ресурсов, перевести их в разряд специальных характеристик ресурсов, обеспечить соответствующую интерпретацию и поддержку. В результате из средства упоминания ресурсов можно получить средство **интеграции** ресурсов. Это даст возможность обеспечить более точный поиск, естественную и удобную навигацию в контексте найденного ресурса, упростить сопровождение и создание ресурсов. Такой механизм аналогичен гипертекстовым связям Web - пространства, с помощью которых можно представлять отношения между ресурсами при Web - доступе к системе. Отношения между типами ресурсов предоставляют средство **структуризации** информационного пространства цифровой библиотеки. С его помощью может быть осуществлена организация ресурсов в иерархические структуры. Например, на основе отношения между организационными единицами можно определить иерархию административного или организационного подчинения. С помощью структурированного рубрикатора можно задать тематическую или видовую рубрикацию определенной пред-

метной области. Можно создавать более сложно связанные образования, например, персональные, коллективные **коллекции** ресурсов.

Разбивая различные метаданные на множества связанных ресурсов, вводя дополнительные типы ресурсов и их связи, мы будем усложнять и усложнять информационную модель библиотеки, смешивать представителей разных предметных областей. С одной стороны, вряд ли можно включить все потенциально необходимые кому-либо типы ресурсов и определить все желательные связи между ними. С другой стороны это может привести к существенным проблемам в реализации и производительности. Чтобы избежать этого, необходимо вложить этот процесс в определенные рамки, обеспечить его технологичность. Необходимо абстрагироваться от деталей отдельных ресурсов и их связей, устранить влияние, которое могут оказать различные программные средства, обслуживающие ресурсы. Основой подобной схемы может послужить концептуальная модель данных, включающая только понятия типов ресурса, свойств (атрибутов) ресурсов и отношений между ними. Отношения должны определяться *между типами* ресурсов. Это связи, характерные для всех ресурсов некоторого типа. Такие отношения можно эффективно использовать и реализовать. Связи, наличие которых определяется не типом ресурса, а содержанием ресурса, его данными, достаточно редки и специфичны. Их поддержку можно обеспечить через программные интерфейсы, описываемые метаданными взаимодействия с ресурсом.

Схема, представленная на рис. 5, позволяет обеспечить гибкость, масштабируемость и настраиваемость системы на всевозможные множества информационных ресурсов, не ограничивая потенциальные требования и пожелания.

Выделение отношений, их типизация дают возможность:

- параметризовать и упростить реализацию поиска и навигации, обеспечить связывание ресурсов в распределенной среде;

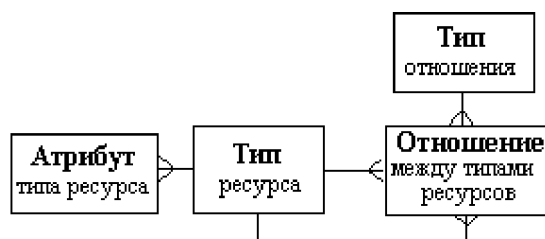


Рис. 5:

- предоставить набор стандартных механизмов связывания ресурсов, например, обеспечить стандартные средства создания тематических коллекций ресурсов, ввода классификаторов и рубрикаторов;
- структурировать предоставляемые услуги поиска информации и навигации в информационном пространстве в зависимости от ориентации релиза системы, предпочтений или бюджетов пользователей (базовый набор услуг, для библиотекарей, платный и т.п.)
- указывать в традиционных форматах метаданных не примерное текстовое описание ресурсов, связанных с рассматриваемыми, а их уникальные идентификаторы - точные ссылки на ресурсы

Весь ли ресурс должен рассматриваться как объект отношений? Все зависит от целей и возможностей системы. Если основная цель системы состоит в том, чтобы обеспечить широкомасштабный, распределенный поиск информации и навигацию по информационному пространству, то объектами отношений должны стать метаданные ресурса, а не он сам непосредственно, причем метаданные, ориентированные на обеспечение поиска и навигации. Сделав поисковые метаданные объектом отношений, мы можем абстрагироваться от, возможно, существенного различия ресурсов и работы с ними, можем существенно упростить и сделать более эффективным распределенный поиск. При этом не существенно, что это за ресурс, какова его специфика, внутренняя структура, правила управления им, важна информация, которая позволяет обеспечить 'разумно-ограниченный' поиск и связать ресурсы в единую информационную сеть. Сам ресурс должен оставаться для поисковой части системы 'черным ящиком', правила управления и взаимодействие с которым определяются остальными метаданными (рис. 6).

Оставшиеся метаданные в основном специфицируют прикладные интерфейсы (SQL подсистемы, API для C++, Java, CORBA компонент, DTD для XML и SGML приложений), посредством которых обязательные или вспомогательные компоненты системы могут осуществлять взаимодействие с ресурсом, обеспечивая соответственно стандартные или расширенные возможности системы по работе с ресурсами определенного типа. К этому виду относятся метаданные, описывающие и/или обеспечивающие:

- (SQL) доступ к информации,
- визуализацию информации в краткой, стандартной и развернутой формах,

- поддержку форматов представления,
- навигацию по структуре информации,
- защиту и управление доступом к ресурсу,
- поддержку актуальности информации,
- редактирование информации и метаинформации,
- управление распределенным поиском и аккумуляцией ответов подчиненных серверов,
- контроль предоставляемых услуг, бюджетов пользователей.

Для обеспечения эффективной реализации распределенного поиска объектом отношений между ресурсами должны быть метаданные поиска, однако, в других случаях этим объектом могут быть и остальные виды метаданных.

В соответствии с использованием сервисами библиотеки отношения между ресурсами, как и элементы метаданных, образуют пересекающиеся группы, которые классифицируются следующим образом:

- Отношения, используемые при **поиске** ресурсов - это отношения, которые могут быть использованы при описании поисковых запросов. Например, автор публикации, участник проекта.
- Отношения, используемые при **просмотре** свойств ресурсов, то есть при предоставлении пользователям информации о ресурсах, их связях с другими ресурсами, для демонстрации информационного контекста ресурса, в котором он находится.
- Отношения, необходимые для **управления операциями** над ресурсом. Сюда относятся отношения, показывающие, с какими ресурсами может быть установлено соединение, обеспечивающее эффективное выполнение операций поиска над рассматриваемым типом ресурса, отношения, определяющие права доступа для поддерживаемых категорий пользователей и т.п.

Другой вид классификации отношений может быть связан с потенциальной частотой их употребления. Например, можно выделить **основные** и **вспомогательные** отношения. Смысл такого разбиения состоит в выделении отношений, которые будут использоваться в большинстве случаев, и отношений, которые будут применяться значительно реже или вообще предназначены для специалистов определенной области знаний (например,

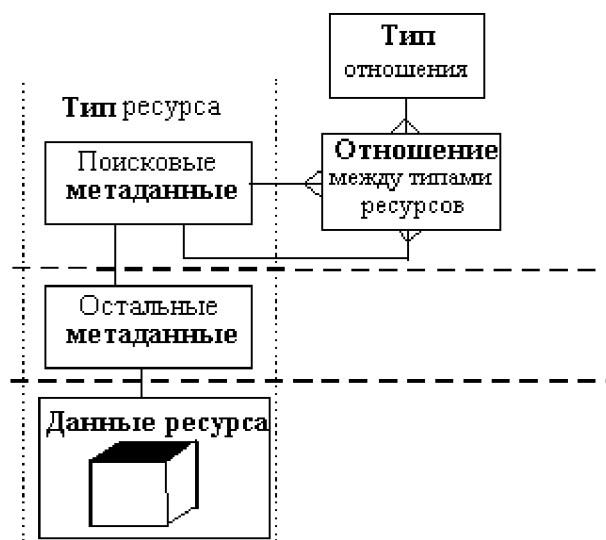


Рис. 6:

библиотекарей). Тогда для первых можно обеспечить эффективную реализацию, позволив определять их только при генерации или настройке системы, а для вторых можно предоставить возможность вводить новые типы отношений, удалять старые, но за счет некоторого понижения производительности при работе с ними. Например, в традиционной цифровой библиотеке отношение *редактор* между *персоной* и *публикацией* скорее всего следует отнести ко второму типу, а отношения *автор* и *работает в* - к первому

3 Архитектура системы ИСИР РАН

Система ИСИР основывается на следующих тезисах вышеизложенной концепции цифровых библиотек:

- в цифровой библиотеке логической единицей хранения является ресурс;
- ресурсы могут быть как информационными, так и программными или вычислительными;
- ресурс имеет две части - метаданные и содержание;
- предоставляется возможность взаимодействовать с непосредственно ресурсами (точнее, с объектами ресурсов) в соответствии с открытыми интерфейсами, поддерживаемыми им;
- уникальный идентификатор ресурса обеспечивает идентификацию ресурса в соответствующей среде;
- каждый ресурс относится к некоторому типу ресурсов, определяющему его свойства;
- элементом метаданных является пара (атрибут, значение), описывающая некоторое свойство ресурса;
- значения атрибутов могут иметь сложную организацию;

- метаданные ресурса подразделяются на метаданные, представляющие внутренне присущие ресурсу свойства (метаданные поиска, просмотра), и метаданные, обеспечивающие взаимодействие с ресурсом и управление службами библиотеки при работе с ним;
- метаданные ресурса описывают сведения свойственные только ему и ни в коей мере не представляют его взаимоотношения с другими ресурсами;
- взаимосвязи ресурсов задаются отдельно от метаданных и определяются понятием отношения между ресурсами;
- отношения устанавливаются между типами ресурсов, а не между ресурсами непосредственно, и могут относиться к некоторому определяемому типу отношений;
- объектами отношений между ресурсами в распределенной среде являются только метаданные поиска и навигации;
- связи, наличие которых определяется не типом ресурса, а содержанием ресурса, его данными реализуются через программные интерфейсы, описываемые метаданными взаимодействия с ресурсом.

Система должна поддерживать следующую функциональность:

- загрузку метаданных в известных форматах;
- выгрузку метаданных в ряд форматов;
- поиск ресурсов по метаданным и отношениям между ресурсами;
- навигацию в пространстве ресурсов;
- редактирование свойств ресурсов и, возможно, содержания ресурсов;

- управление доступом к ресурсам;
- распределение ресурсов в сети;
- обеспечение актуальности ресурсов;
- поддержку взаимодействия с ресурсами.

В системе должны использоваться следующие виды элементов метаданных:

- метаданные поиска;
- метаданные идентификации ресурсов;
- метаданные просмотра для различных форм визуализации;
- метаданные контроля редактирования;
- метаданные управления доступом;
- метаданные поддержки актуальности ресурсов;
- метаданные обеспечения взаимодействия с ресурсом (стандартная поддержка ряда форматов данных, программных интерфейсов).

Метаданные поиска, просмотра, идентификации, актуализации специфичны для ресурсов, то есть каждый из них обладает собственным набором значений для этих элементов метаданных. Метаданные контроля доступа, редактирования, управления поиском, просмотром, обеспечения взаимодействия (то есть параметры настройки стандартных механизмов системы) ассоциируются с типами ресурсов. Для большинства элементов последней категории в системе определены наборы допустимых значений, семантика которых поддерживается системой.

Функционирование системы обеспечивается рядом системных программных компонент (группами базовых объектов), выполняющих групповые операции над ресурсами, такие как загрузка, выгрузка, поиск, вывод результатов поиска, активирующих и поддерживающих деятельность объектов ресурсов в процессе обслуживания соответствующих им ресурсов. Все объекты ресурсов относятся к классам объектов, соответствующим типам ресурсов. Все классы объектов ресурсов являются наследниками предопределенного класса `Resource`, виртуальные методы которого обеспечивают поддержку стандартных служб и механизмов системы.

Механизмы загрузки/выгрузки ресурсов ориентированы на ряд стандартных форматов обмена метаданными (`DC+HTML`, `DC+RDF`, `RDF`, `XML`) и настраиваются на тип ресурса по его описанию. Они имеют выражение в виде системной компоненты загрузки/выгрузки, используемой при массовой загрузке/выгрузке ресурсов, и в виде стандартных методов объектов ресурсов, применяемых при работе с отдельными ресурсами, например, при формировании `META` - тегов `HTML` - страниц ресурсов.

Операции поиска зависят от типов ресурсов, отношений между ними и не привязаны к конкретным экземплярам ресурсов, поэтому также реализуются системной компонентой поиска. Эта компонента использует описания ресурсов, их отношений, метаданные управления поиском при преобразовании поисковых запросов в `SQL` - запросы.

Просмотр свойств ресурсов реализуется по-разному в зависимости от формы представления ресурсов. Краткая форма, в которой краткие сведения о ресурсе являются ссылкой на стандартную форму представления сведений

о ресурсе, используется системной компонентой поиска при выводе результатов поисковых запросов. Стандартная и развернутая формы просмотра и навигации формируются методами объектов ресурсов, настраиваемыми на тот или иной вид визуализации в соответствии с описанием ресурса, его метаданными управления доступом, просмотром и навигацией. В случае `Web` - интерфейса связи ресурсов реализуются с помощью гипертекстовых ссылок.

Редактирование свойств и, возможно, содержания ресурса обеспечивается методами объектов ресурсов, руководствующимися описанием ресурса, метаданными управления доступом, редактированием. Поддержка функциональности свойственной отдельным ресурсам, их механизмов взаимодействия может осуществляться объектами ресурсов либо с помощью совокупности классов объектов для поддерживаемых системой форматов представления данных (`TXT`, `HTML`, `RDF`, `XML`, `binary`), либо посредством стандартных для системы программных интерфейсов (`API` для `Java` и `C/C++`).

3.1 Мета модель

В основе реализации системы ИСИР лежит модель, исходящая из того, что ресурс характеризуется набором присущих ему атрибутов и совокупностью взаимоотношений с другими ресурсами. Это позволяет воспользоваться хорошо себя зарекомендовавшими принципами информационного моделирования.

В любом информационном приложении в его информационной модели всегда можно выделить ряд сущностей или их совокупностей, которые представляют информационную составляющую предметной области, вокруг которых сконцентрирована логика приложения, набор предоставляемых сервисов, которые составляют скелет информационной модели. Такие сущности иногда называют бизнес сущностями или ключевыми сущностями. Это сущности уровня модели данных бизнеса (`business data model`), которая отражает требования бизнес данных. Среди них нет сущностей системной модели данных (`system data model`), детализирующих описание предметной области и ориентирующихся на поддержку приложений, работающих с данными. Эти сущности усложняют описание модели данных, вносят в схему массу зависящих от приложений деталей, деуалирует свойства данных предметной области.

Ресурсы можно рассматривать как сущности уровня модели данных бизнеса. Атрибуты ресурсов, в общем имеющие сложное строение, описывают свойства ресурсов и детали предметной области, для выражения которых в информационном моделировании используются сущности системной модели данных. Следуя этому, мы предлагаем описывать предметную область библиотеки, думая только о том, что характеризует ресурс, что должно использоваться при работе с ресурсом. В чем состоит работа с ресурсом, как он будет использоваться, фиксируется самим понятием цифровой библиотеки. В ее базовый набор функциональности как приложения входят такие операции, как хранение, поиск, просмотр, извлечение, загрузка и выгрузка, возможно, редактирование.

Уменьшая возможности по определению свойств данных, навязывая в связи с этим некоторые ограничения целостности, предоставляя предопределенные высокоуровневые понятия и свойства, мы упрощаем и облегчаем описание предметной области библиотеки. Зная, какого рода функциональностью должно обладать прило-

жение, можно автоматизировать введение в модель данных сущностей системной модели, необходимых для поддержки приложения. Следовательно, можно обеспечить автоматическое построение корректной схемы базы данных и приложения, работающего с ней, по высокоуровневому описанию предметной области библиотеки. Чтобы предоставить некоторую свободу управления функциональностью библиотеки, можно осуществить параметризацию функций библиотеки.

Система ИСИР реализует этот подход, позволяющий автоматизировать создание цифровых библиотек. Информация о ресурсах системы, их атрибутах, отношениях между ресурсами сохраняется в так называемом метарепозитарии системы. Это отдельная схема (подсхема) базы данных, содержащая описание информационной модели предметной области, параметров настройки стандартных функций библиотеки. В метарепозитарии может храниться описание как одной, так и нескольких предметных областей. Схема метарепозитария показана на рис. 7.

По информации из метарепозитария осуществляется генерация схемы базы данных цифровой библиотеки и первичное заполнение служебной базы данных, в которой хранятся данные, обеспечивающие поддержку стандартных функций библиотеки, динамически определяемые отношения между ресурсами и т.п. Ресурсы предметной области, их связи поддерживаются отдельной подсхемой - БД ресурсов библиотеки (рис. 8).

3.2 Поиск

Для обеспечения поиска ресурсов и обмена данными между узлами распределенной цифровой библиотеки в системе ИСИР поддерживаются два вида языков запросов. Языки не привязаны к конкретному набору типов ресурсов и отношений между ними. Настройка языков на ту или иную предметную область, конвертирование поисковых запросов в команды языка SQL осуществляется на основе описания предметной области, хранящегося в метарепозитарии.

Один язык запросов (QL) предназначен для внутренних служб системы и используется ею в процессе взаимодействия серверов системы при обмене распределенными данными - их актуализации, кэшировании и т.п. Это некоторая форма стандартного языка SQL, в которой вместо понятий таблиц, ключей и элементарных данных используются понятия ресурса, отношения между ресурсами, сложных типов данных и способов выделения элементов сложных типов данных.

Второй вид запросов (SL) используется при поиске ресурсов и применяется всеми поисковыми интерфейсами системы (Web, Z39.50 [4], LDAP [5, 6]) при формулировке поисковых запросов к библиотеке. В этом языке выделены понятия двух видов подвыражений: F-выражения и J-выражения. Первые задают ограничения на атрибуты ресурсов (фильтрует ресурсы некоторого типа), а вторые специфицируют условия на наличие отношений между ресурсами (соединяет ресурсы).

- F-выражение позволяют описать условия, которым должны удовлетворять атрибуты ресурса определенного типа, и объединить эти условия операциями OR, AND и NOT. Условия позволяют задать следующие ограничения на значения атрибутов ресурсов:

– присутствует или нет;

- попадает или нет в некоторый интервал, диапазон;
- равен или не равен некоторому значению;
- является ли словом (цепочкой символов), удовлетворяющей заданному шаблону;
- содержит ли слова, расположенные на заданном расстоянии.

- J-выражение - это выражение, содержащие условия, объединяемые операциями OR, AND и AND-NOT, требующие наличия или отсутствия отношений между ресурсом, представляемым F-выражением, и другими ресурсами, указываемыми J- или F-выражениями. Чтобы иметь возможность несколько раз сослаться на ресурс, удовлетворяющий некоторому выражению, при указании его связей с другими ресурсами, с этим выражением можно связать идентификатор и использовать его в дальнейшем в формулировке запроса.

Строгость базовой формы языка обеспечивают эффективную обработку запросов и позволяет создать единый механизм доступа к ресурсам, используемый разнообразными поисковыми протоколами системы.

Наряду со строгостью языка, требующей указания того, какой атрибут должен принимать то или иное значение, обязывающей явно разделять ограничения на атрибуты и отношения, в этом поисковом языке имеется ряд правил, упрощающих формулировку запросов и позволяющих охватить большинство возможностей, предлагаемых современными поисковыми системами. Например, в F-выражении для одного ресурса можно задать условие на значение атрибута другого ресурса, указывая имя отношения и, возможно, имя типа ресурса в описании квалификационного имени атрибута. Если в ограничении на значение атрибута опущено его имя, то такое условие считается истинным, если условие удовлетворяет хотя бы один атрибут ресурса. Такое же правило действует в случае сложно структурированных атрибутов. В поисковом выражении, состоящем только из одного F-выражения, можно опустить имя искомого ресурса. Это означает, что необходимо найти ресурсы любых типов, среди атрибутов которых имеется атрибут, удовлетворяющий этому F-выражению.

Метаописание цифровой библиотеки, сохраняемое в метарепозитарии, позволяет управлять допустимостью тех или иных правил упрощения, определять какие из отношений, типов ресурсов, их атрибутов ресурсов должны использоваться по умолчанию. Правила упрощения дают пользователям возможность достаточно просто формулировать сложные поисковые запросы, хотя и язык SL не предназначен для непосредственного использования пользователями, поскольку для этого имеются специальные визуальные средства, упрощающие создание поисковых запросов различных уровней сложности.

Система использует язык запросов SL как внутренний формат представления поисковых запросов, обеспечивающий управляемый перевод поисковых запросов разнообразных протоколов (Web, Z39.50, LDAP) в последовательность SQL запросов. Такая форма реализации, кроме унификации поддержки разных видов поиска и серверов баз данных, необходима для того, чтобы обеспечить приемлемое преобразование поисковых запросов. Это позволяет избежать выполнения большого количества отдельных SQL запросов, порождающих множества,

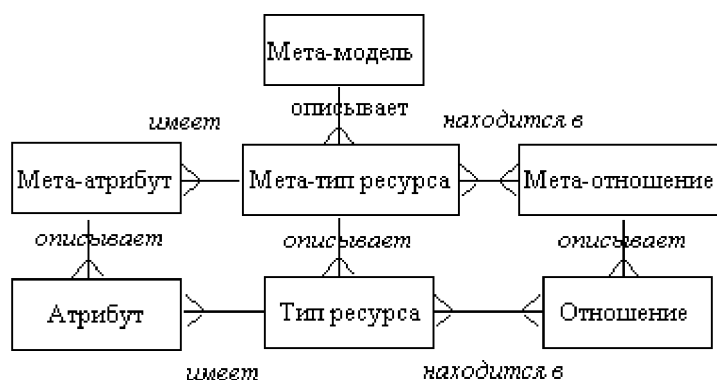


Рис. 7:

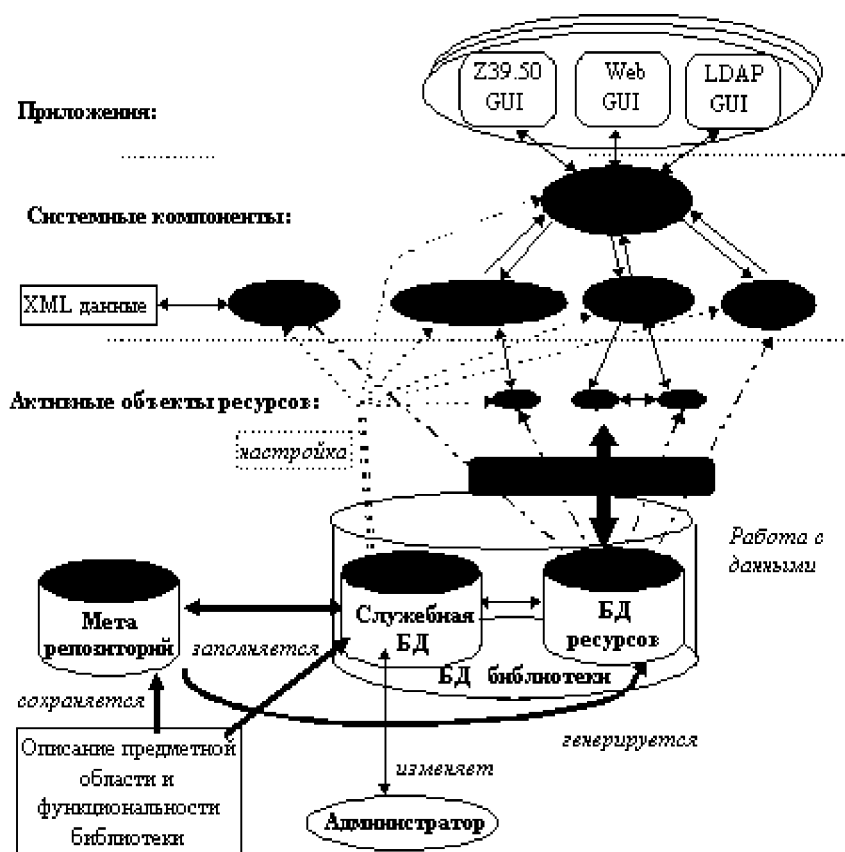


Рис. 8:

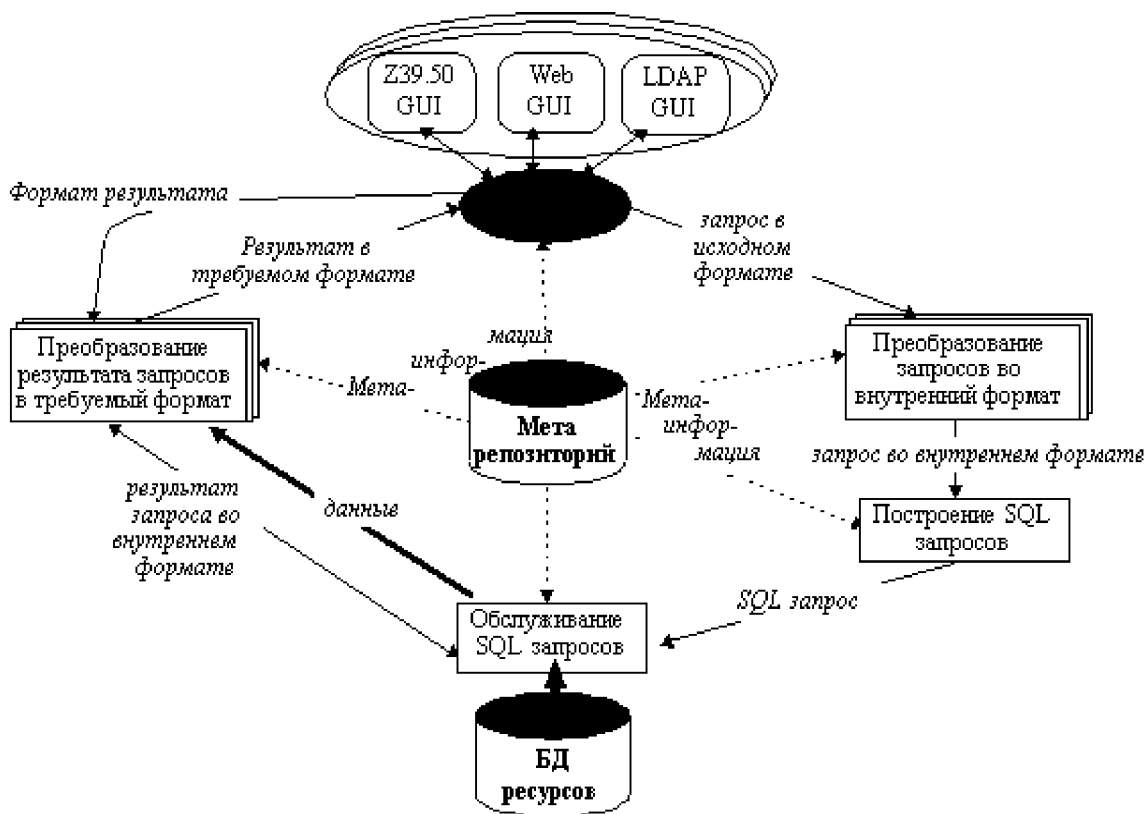


Рис. 9:

сопоставимые по мощности со всем множеством ресурсов системы, избавиться от сохранения промежуточных результатов. Преобразование исходных поисковых запросов (рис. 9) во внутренний формат, генерация SQL запросов производится на основе метайнформации из метарепоzitория, описывающей предметную область и характеризующей функциональность поискового протокола.

Формирование результатов запроса также осуществляется на основе внутреннего представления. Это позволяет избежать накладные расходы при работе с ресурсами, обладающими развернутой структурой, при подготовке сложных отчетов. Дает возможность иметь гибкие средства для поддержания разнообразных форматов результатов запросов, обеспечить сохранение результатов запросов для протоколов подобных Z39.50.

3.3 Распределенность

Система серверов, составляющих интегрированную систему РАН и распределенных между организациями РАН, должна обеспечивать различные уровни работы с ресурсами (рис. 10), в зависимости от возможностей и политики организации, сопровождающей сервер. Одни сервера должны хранить только ресурсы (скорее всего, только собственные), вторые должны обеспечивать только поиск по метаданным, третьи должны быть полнофункциональными:

- обеспечивать хранение и доступ к ресурсам;
- осуществлять поиск в своем множестве поисковых метаданных;
- перенаправлять запросы другим поисковым серверам системы.

Необходимо, чтобы сервер организации имел возможность функционировать самостоятельно (независимо от остальных) как информационная система этой организации. В каждом узле системы могут использоваться различные технологии, применение релизов ИСИР обязательным не является. Однако, требуется, чтобы каждый узел обеспечивал единообразный взгляд на информацию в терминах 'ресурсов-отношений', реализуя интерфейсы системы ИСИР, основанные на открытых стандартах.

В системе ИСИР ключевыми понятиями организации распределенного хранения и доступа к интегрированной информации являются: ресурс, отношение между ресурсами, поисковые метаданные и уникальный идентификатор ресурса.

Уникальный идентификатор ресурса, который можно рассматривать как аналог ISSN или ISBN, позволяет добиться относительной независимости узлов от центральной службы. Он следует принципам *универсального имени ресурса* (URN - uniform resource name), введенного IETF (Internet Engineering Task Force) в 1994 г [22, 23] и призванного идентифицировать информационные ресурсы вне зависимости от их местоположения в Интернет. URN является постоянным и уникальным идентификатором. URN может быть присвоен только ресурсу, являющемуся достаточно 'стабильным' и имеющему 'существенное' интеллектуальное, информационное содержание'. Будучи сопоставленным ресурсу, он не может быть изменен. Копии ресурса должны использовать один и тот же URN вне зависимости от формы представления информации. 'Тот же' ресурс может получить новый URN,

только если его интеллектуальное содержание изменилось, но, следовательно, это уже 'новый' ресурс. 'Старый' ресурс может быть сохранен и остаться доступным по старому URN. Он может быть удален, но его 'использованный' URN не может быть использован для другого ресурса

Использование общепризнанного стандарта позволяет избавиться от частных схем именования и сопутствующих им проблем переименования, обеспечить интеграцию информации в Интернет и поддержку ее актуальности. Чтобы поддержать существующие общепризнанные и специализированные схемы именования источников информации, URN вводит понятие пространства имен (name space) как именованной совокупности имен, соответствующих некоторой зарегистрированной схеме. Для обеспечения понятия уникального имени используются системы *регистрации* пространства имен, *формирования* и *разрешения* имен. Первая обеспечивает международную регистрацию схем именования и относится к ведению IETF или связанной с ней организацией. Вторая специализируется на конкретной схеме и отвечает за поддержку уникальности имен. Третья должна сообщать по URN сведения об именуемом ресурсе, в которые могут быть включены данные о месте положения ресурса, о способе доступа к нему по сетевым протоколам. В Интернет среде в качестве источников таких сведений рассматривается URL (Uniform Resource Locator) и URC (Uniform Resource Citation). Имеет ряд реализаций систем именования и их Интернет поддержки, например, Handle System [24], DOI [10].

В системе ИСИР каждому серверу назначается собственный сегмент именования в зарегистрированном пространстве имен. В рамках этого сегмента организация отвечает за уникальность имен своих ресурсов. Вышестоящая организация может выделить часть своего сегмента своим подчиненным организациям.

Для обеспечения распределенного поиска и хранения используются следующие соглашения:

- Каждый ресурс обладает уникальным идентификатором.
- Каждому ресурсу сопоставлен один и только один сервер, *отвечающий* за него - его метаданные и содержание. Изменение, редактирование ресурса ведется только этим сервером. Другие сервера могут иметь копии ресурса. Они должны следовать одной из дисциплин поддержки актуальности копии ресурса (временной период, сигнал об изменении). Копии создаются и используются только для обеспечения эффективного доступа к содержанию ресурса.
- Каждый сервер имеет некоторое множество (возможно, пустое) ресурсов, за которые он отвечает. Кроме метаданных собственных ресурсов сервер обязан иметь копии *поисковых* метаданных ресурсов, которые *непосредственно связаны* с его собственными ресурсами, в соответствии с отношениями между типами ресурсов. Например, если публикация, ведомая некоторым сервером, имеет персону соавтора, за которую отвечает другой сервер, то первый сервер обязан иметь копию поисковых метаданных персоны, а второй - копию поисковых метаданных публикации.
- Для организации эффективного поиска и обеспечения отказоустойчивости, особенно, при малопроеиз-

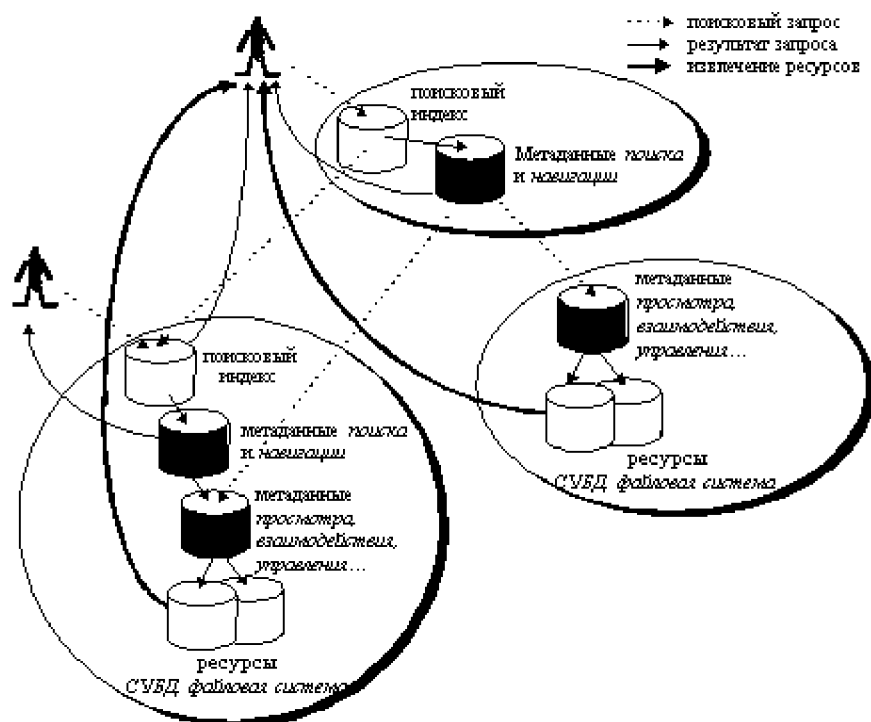


Рис. 10:

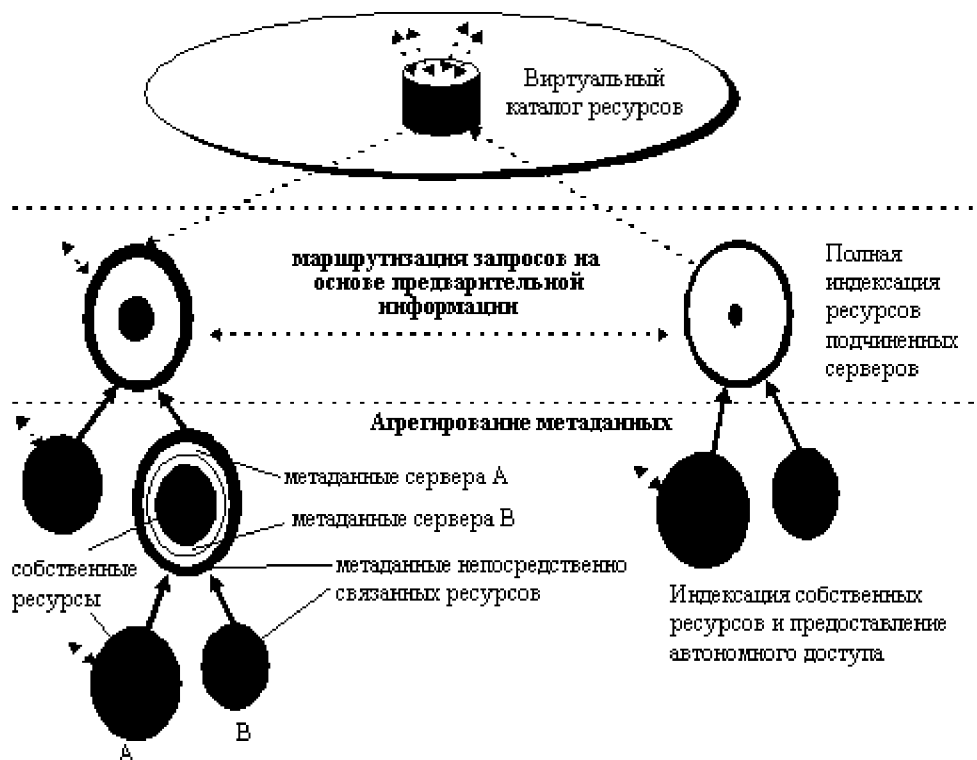


Рис. 11:

водительных сетях и компьютерах, один сервер может аккумулировать поисковые метаданные подчиненных ему серверов. Это обеспечивает более эффективную обработку поисковых запросов и не требует больших объемов памяти для хранения информации. Такой способ индексирования информации при аккумулировании метаданных на основе некоторого критерия дает возможность создавать коллекции информации, аналогичные [26].

- Сервера могут образовывать сегменты, в которых обеспечивается 'широковещательная' рассылка поисковых запросов. Ответом системы на некоторый запрос является сумма ответов об собственных ресурсах отдельных серверов системы. Технологии типа forward knowledge & query routing (маршрутизация запросов на основе предварительной информации) [25] позволяют оптимизировать прохождение распределенных запросов при такой организации взаимодействия.

В случае ИСИР РАН структура серверов соответствует сложившейся организационной структуре РАН (рис. 11), то есть информационные ресурсы должны быть иерархически каталогизированы в соответствии с организационной структурой РАН. Каждый узел иерархии может иметь ограниченный или полнофункциональный вариант системы. До некоторого уровня этой иерархии осуществляется индексация всех информационных ресурсов, соответствующих данному уровню. Выше этого уровня обеспечивается 'широковещательная' рассылка поисковых запросов.

3.4 Текущее состояние

В настоящее время имеется первая версия системы ИСИР РАН, к которой можно обратиться по адресу <http://isir.ras.ru/>. Система поддерживает ресурсы: персона, организация, подразделение, публикация и проект. Реализована совокупность стандартных типов, упрощающих описание цифровой библиотеки. Система является многоязычной, но на данный момент имеются данные только на русском и английском языках. Для поддержки иерархические связей, рубрикации ресурсов в системе реализованы специальные механизмы навигации, учитывающие специфику соответствующих типов ресурсов и настраиваемые по их описаниям.

Используя Web-интерфейс, можно искать ресурсы по значениям их атрибутов, редактировать атрибуты. Можно осуществлять навигацию в пространстве ресурсов, просматривая сведения о них, загружая их содержание. Обеспечивается четыре уровня доступа к ресурсам. Метаданные ресурсов содержатся в реляционной БД; содержание ресурсов, если оно имеется, хранится либо в реляционной БД, либо в файловой системе в зависимости от вида ресурса. Ресурс может иметь несколько экземпляров содержания, возможно, представленных в разных форматах.

Имеются реализации системы для платформ UNIX и MS Windows/NT, использующие RDBMS Oracle и Oracle Web-сервер.

Выполнены пилотные реализации доступа по протоколам Z39.50 и LDAP, в которых были отработаны механизмы трансляции поисковых запросов во внутреннем формате в язык SQL и организации распределенного взаимодействия и поиска. Создан загрузчик метаданных формата Dublin Core, представленных в XML [11, 12] синтаксисе методики RDF [13, 14].

Библиография

- [1] Kahn, R. and R. Wilensky. *A Framework for Distributed Digital Object Services*, May 1995; <http://www.cnri.reston.va.us/k-w.html>.
- [2] William Y. Arms. *Key Concepts in the Architecture of the Digital Library*, D-Lib Magazine, July 1995; <http://www.cnri.reston.va.us/k-w.html>.
- [3] Stuart Weibel and Eric Miller. *Dublin Core Metadata*; http://purl.org/metadata/dublin_core
- [4] Z39.50 Maintenance Agency; <http://lcweb.loc.gov/z3950/agency>
- [5] M. Wahl, T. Howes and S. Kille. *Lightweight Directory Access Protocol (v3)*, RFC 2252; <http://ds.internic.net/rfc/rfc2251.txt>
- [6] Wahl, M., Coulbeck, A., Howes, T., and S. Kille. *Lightweight Directory Access Protocol (v3): Attribute Syntax Definitions*, RFC 2252; <http://ds.internic.net/rfc/rfc2252.txt>
- [7] *DC Elements, Reference Version 1.0*; http://purl.oclc.org/metadata/dublin_core_elements
- [8] Бездушный А.Н. *Подход к созданию интегрированных цифровых библиотечных систем*, ОСМО ВЦ РАН, ТО-04-1999-4
- [9] Бездушный А.Н. *Обзор форматов метаданных*, ОСМО ВЦ РАН, ТО-02-1999-5
- [10] *International DOI Foundation web site*; <http://www.doi.org>
- [11] Tim Bray, Jean Paol and C. M. Sperberg-McQueen. *Extensible Markup Language (XML) 1.0*; <http://www.w3.org/TR/REC-xml>
- [12] Tim Bray, Dave Hollander and Andrew Layman. *Namespaces in XML*; <http://www.w3.org/TR/REC-xml-names>
- [13] Dempsey, Lorcan, Rachel Heery, Martin Hamilton, Debra Hiom, Jon Knight, Traugott Koch, Marianne Peereboom, and Andy Powell. *A Review of Metadata: A Survey of Current Resource Description Formats. Work Package 3 of Telematics for Research project DESIRE (RE 1004). Version 1.0*; <http://www.ukoln.ac.uk/metadata/DESIRE/overview/>
- [14] Lassila, Ora and Ralph R. Swick. *Resource Description Framework (RDF) Model and Syntax*, W3C Working Draft, <http://www.w3.org/TR/REC-rdf-syntax>
- [15] Dan Brickley, R.V. Guha and Andrew Layman. *Resource Description Framework, (RDF) Schemas*, W3C Working Draft; <http://www.w3.org/TR/WD-rdf-schema>,
- [16] *Harvest system web site*; <http://www.tardis.ed.ac.uk/harvest/>
- [17] Luis Gravano, Kevin Chang, Hector Garcia-Molina, Carl Lagoze, Andreas Paepcke. *STARTS: Stanford Protocol Proposal for Internet Retrieval and Search*; <http://www-db.stanford.edu/gravano/starts.html>

- [18] Y. Goland , E. Whitehead, A. Faizi, S. Carter, D. Jensen. *HTTP Extensions for Distributed Authoring - WEBDAV*, RFC 2518;
<http://ds.internic.net/rfc/rfc2518.txt>
- [19] Saveen Reddy, Dale Lowry, Surendra Reddy, Rick Henderson, Jim Davis, Alan Babich. *DAV Searching & Locating*;
<http://www.webdav.org/dasl/protocol/draft-dasl-protocol-00.html>
- [20] Sandra Payette, Christophe Blanchi, Carl Lagoze, Edward A. Overly. *Interoperability for Digital Objects and Repositories*, D-Lib Magazine, May 1999, Volume 5 Issue 5, ISSN 1082-9873;
<http://www.dlib.org/>
- [21] *The World Wide Web Consortium Home Page*,
<http://www.w3.org/>
- [22] Sollins, K., and L. Masinter. *Functional Requirements for Uniform Resource Names*, RFC 1737, MIT/LCS, Xerox Corporation, December 1994.
- [23] Berners-Lee, T., Masinter, L., and M. McCahill, Editors. *Uniform Resource Locators (URL)*, RFC 1738, CERN, Xerox Corporation, University of Minnesota, December 1994.
- [24] *The Handle System Home Page*,
<http://www.handle.net/>
- [25] John Kirriemuir, Dan Brickley, Susan Welsh, Jon Knight, and Martin Hamilton. *Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach*;
<http://www.dlib.org/dlib/january98/01kirriemuir.html>
- [26] Carl Lagoze, David Fielding. *Defining Collections in Distributed Digital Libraries*