

# Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек

*Бездушный А.Н., e-mail: bezdushn.ccas.ru,*

*Жижченко А.Б., e-mail: abz@ipsunras.ru,*

*Кулагин М.В., e-mail: ql@ras.ru,*

*Серебряков В.А., e-mail: serebr.ccas.ru*

## Аннотация

В статье рассматриваются проблемы разработки цифровых (электронных) библиотек. Проведен анализ свойств этого класса распределенных информационных систем. Основное внимание уделено формализации свойств объектов хранения и операций, что дает возможность построить формальную модель таких систем, на основе которой может быть разработан инструментарий для их разработки. Систему, основанную на этом подходе, мы называем интегрированной системой информационных ресурсов (ИСИР). На принципах ИСИР реализована информационно-справочная система Российской академии наук (ИСИР РАН).

Статья организована следующим образом. Вначале рассматриваются цели работы. Затем дается определение цифровой библиотеки (в понимании авторов), обсуждаются основные понятия цифровой библиотеки и ИСИР рассматривается как цифровая библиотека. Вводится понятие ресурса как основного объекта обработки. Рассматриваются метадаанные и их роль в системе цифровых библиотек, роль отношений в организации данных и работе с ними. Далее рассматривается архитектура и функциональность ИСИР, в частности метамодель системы, вопросы поиска, реализации распределенности. Кратко рассматривается текущая реализация ИСИР РАН.

## Введение

В последнее время все больше внимания уделяется проблемам создания распределенных информационных систем, обладающих рядом новых черт, которые позволяют выделить их в особый класс, получивший название *цифровых (электронных) библиотек* (ЦБ). Точное определение цифровой библиотеки дать трудно, да и почти невозможно. Однако можно сказать, что непосредственными предшественниками цифровых библиотек являются традиционные информационные системы, в частности распределенные, и традиционные библиотечные системы, также предоставляющие электронный сервис. Можно сказать, что ЦБ - это качественный скачок в области информационных технологий, связанный с интенсивным развитием Интернет-технологий, международных стандартов на программное обеспечение, возможностей каналов связи.

Понятно, что в каждом конкретном случае информационная система, причисляемая к классу ЦБ, обладает определенным набором характеристик. Чтобы ограничить предмет обсуждения, в статье дается более узкое определение ЦБ, что позволяет сосредоточить внимание на определенных аспектах их разработки. Следует подчеркнуть, что это определение не претендует на всеобщность – речь идет лишь о более строгом определении некоторого подкласса таких систем. Из такого более строгого, но и более ограниченного понимания предмета исследования можно вывести определение функциональности ЦБ, ее структуру и задачи разработки.

В статье описывается подход к разработке инструментария для создания ЦБ рассмотренного класса. Работа выполняется в рамках Межведомственной программы Миннауки РФ «Создание национальной сети компьютерных телекоммуникаций для науки и высшей школы», проект 4.28 «Создание Интегрированная Система Информационных Ресурсов РАН» и проекта РФФИ 99-07-90139 «Интегрированная информационная система "Научный Институт" РАН». В рамках этих проектов ведется разработка среды создания ЦБ рассматриваемого класса и одновременно реализация конкретных ЦБ. Одним из примеров может служить Интегрированная Система Информационных Ресурсов Российской академии наук (ИСИР РАН), вкратце рассматриваемая в статье.

Таким образом в статье аббревиатура «ИСИР» употребляется для обозначения различных, хотя и близко связанных, понятий:

- определенного класса цифровых библиотек;
- технологии создания цифровых библиотек указанного класса;
- инструментальной поддержки этой технологии;
- конкретных цифровых библиотек, созданных по этой технологии.

Как правило, сложностей с употреблением аббревиатуры «ИСИР» не возникает, однако, при необходимости мы даем соответствующие пояснения.

## 1. Основные понятия цифровой библиотеки

Как уже было сказано, ЦБ в общем случае представляют собой весьма широкий класс распределенных информационных систем. Для того, чтобы придать определенность дальнейшему изложению, введем более узкое определение. Концептуальная модель системы не нарушает основополагающих принципов ЦБ, изложенные в работах [1] и [2], в которых были определены основные понятия ЦБ. 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 необходимые *программные* сервисы или *вычислительные* ресурсы.

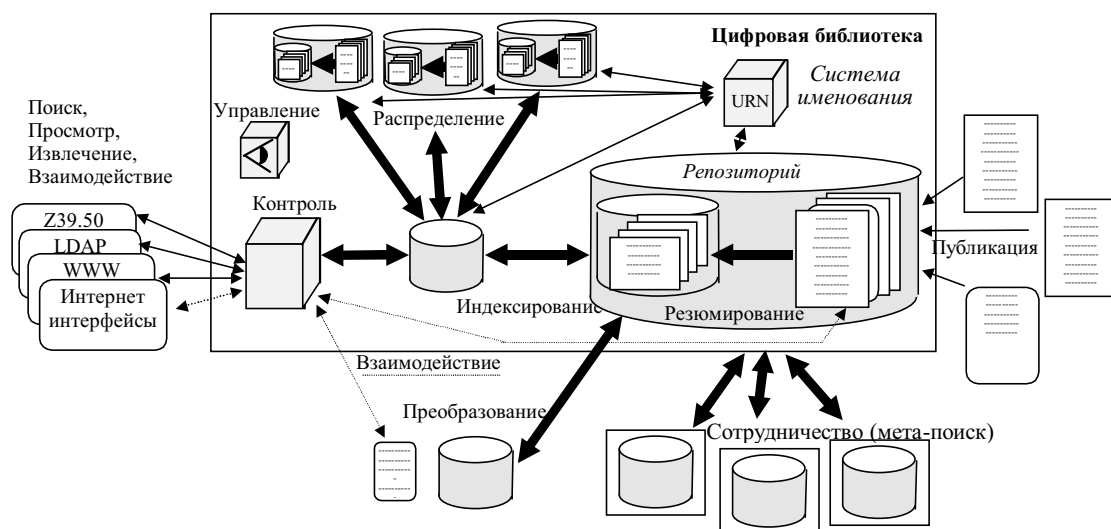
### 1.1. Функциональность цифровой библиотеки

Логическую единицу хранения в ЦБ будем называть *ресурсом* или объектом. Ресурс имеет две части - *метаданные* и, возможно, *данные* (*содержание ресурса*). Метаданные - это сведения о ресурсе, на основе которых осуществляется управление ресурсом, его поиск и взаимодействие с ним. К метаданным относятся и *уникальный идентификатор* ресурса, обеспечивающий идентификацию ресурса в соответствующей локальной или распределенной сетевой среде. Интерпретация содержания ресурса осуществляется на основе его типа или интерфейсов, описываемых в метаданных. Ресурсы библиотеки *не однородны*. Это проявляется в различии их свойства предоставляемых услуг. Каждый ресурс относится к некоторому типу, который называется *типом ресурса* и определяет:

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

Выбор набора типов ресурсов зависит от ориентации ЦБ, ее предметной области. Предполагается, что ЦБ поддерживает следующую функциональность:

- доступ к ресурсам:  
поиск ресурсов на основе метаданных, в частности полученных в результате обработки полнотекстовых документов, навигацию в пространстве связанных ресурсов, просмотр свойств ресурсов;
- администрирование ресурсами:  
загрузку метаданных и содержания ресурсов из разнообразных форматов, выгрузку метаданных и содержания ресурсов в разнообразные форматы, автоматизированную подготовку метаданных, публикацию, удаление ресурсов, редактирование их свойств и, возможно, самих ресурсов, индексирование метаданных, поддержку средств идентификации ресурсов, поддержку актуальности ресурсов, управление доступом к ресурсам и аудит доступа;
- взаимодействие с ресурсами:  
извлечение ресурсов, предоставление содержания ресурса, поддержку и обеспечение функциональности ресурсов на основе открытых интерфейсов, обработку (возможно, аналитическую) информационного наполнения;
- поддержку распределенности ресурсов:  
аккумуляцию и кэширование метаданных, перенаправление, широковещательную рассылку поисковых запросов, интеграцию и идентификацию внешних ресурсов, актуализацию информации о внешних ресурсах;
- мета-поиск (поиск в среде цифровых библиотек и поисковых систем иной структуры);
- преобразование поисковых запросов;
- интеграцию результатов ответов на поисковые запросы.



**Рис. 1. Задачи, решаемые ЦБ, и их взаимодействие**

На рис. 1 изображена возможная схема обеспечения требуемой функциональности ЦБ. Здесь выделены следующие группы задач, которые должна решать реализация ЦБ:

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

## 1.2. Метаданные

*Метаданные* - это сведения о ресурсах, свойства ресурсов. 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 свойств ресурса, а значение задает это свойство. Значения атрибутов могут иметь *сложную* организацию, соответствующую структуре свойства - его подразделению на части, взаимному влиянию частей.

В зависимости от ориентации, уровня абстракции ЦБ одни и те же данные 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 метадаанных службами ЦБ элементы метадаанных образуют пересекающиеся группы, которые можно классифицировать следующим образом:

- Метаданные для **поиска** представляют те элементы метадаанных, которые наиболее часто будут использоваться пользователями при формировании поисковых запросов или в процессе навигации в информационном пространстве библиотеки. Например, название, ключевые слова описания, дата, данные для полнотекстового индекса.
- 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 метадаанных.
- 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 удалять 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 метадаанными.

### 1.3. Отношения между ресурсами и интеграция ресурсов

Ресурсы тесно **взаимосвязаны** друг с другом, что обусловлено как взаимными характеристиками (*персона* является автором *публикации*, которая подготовлена в рамках *проекта*, выполняемого в *организации*), так и взаимными действиями (*программа* транслируется *компилятором*, работает под управлением *ОС* и обращается к *процессам*). Связи, как и ресурсы, не однородны и типизированы. Необходимо выделить из метадаанных описание взаимосвязи (**отношений**) ресурсов, перевести их в разряд специальных характеристик ресурсов, обеспечить соответствующую интерпретацию и поддержку, получить средство **интеграции** ресурсов. В данном случае ИСИР рассматривается как средство интеграции подсистем, т.е. установление *идентичности* одних и тех же объектов в различных подсистемах, возможность навигации по информации о ресурсах в соответствии с установленными отношениями, возможность непосредственного доступа к соответствующей подсистеме через информацию, загруженную в ИСИР.

## 2. Архитектура и функциональность предлагаемой реализации ИСИР

Таким образом, в соответствии с проведенным выше анализом, мы ограничимся в дальнейшем рассмотрением определенного класса ЦБ, которые мы будем называть «Интегрированными Системами Информационных Ресурсов». Теперь мы рассмотрим формальную модель систем этого класса, которая

может послужить основой во-первых, реализации конкретной ИСИР как цифровой библиотеки, и во-вторых, основой для разработки инструментария для создания таких систем. Мета-модель ИСИР служит формальной основой для реализации конкретной ЦБ. Реализованная на основе метамодели конкретная ЦБ представляет собой как-бы «заготовку» без данных. Вопрос о том, какими средствами осуществляется реальное заполнение, в статье не рассматривается.

## 2.1. Метарепозитарий

Информация о ресурсах системы, их атрибутах, отношениях между ресурсами, составляющих модель конкретной реализации ИСИР, сохраняется в так называемом метарепозитарии системы. Это отдельная база данных, хранящая описание информационной модели предметной области, параметров настройки стандартных операций библиотеки. В метарепозитарии может храниться описание предметной области. Схема метарепозитария показана на рис. 2.

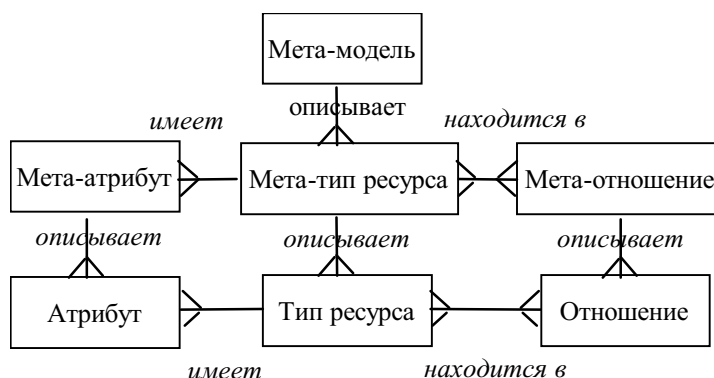


Рис. 2. Схема метарепозитария

Информация в ЦБ хранится в базе данных - репозитарии ЦБ. По информации из метарепозитария осуществляется генерация схемы этой базы данных и ее первичное заполнение, в которое включаются данные, обеспечивающие поддержку стандартных операций библиотеки (Рис. 3)

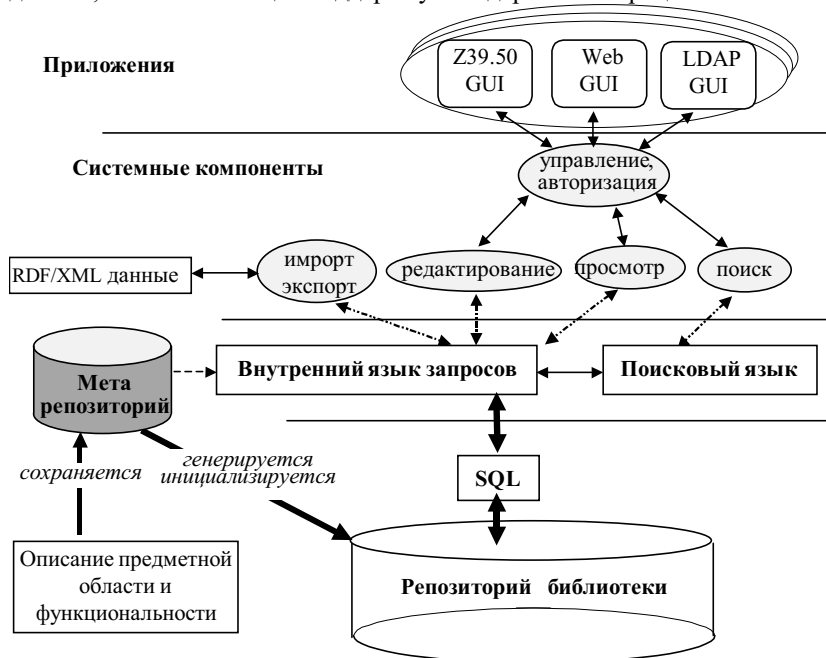


Рис. 3. Функциональная схема ЦБ ИСИР

## 2.2. Языки запросов

Для обеспечения поиска ресурсов и обмена данными между узлами распределенной цифровой библиотеки в 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 запросов СУБД (SQL, OQL) осуществляется на основе описания предметной области, хранящегося в метарепозитории.

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 запросов СУБД, 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[3], 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 ANDNOT, 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 обеспечивает 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 условию 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 метарепозитории, позволяет управлять допустимостью тех или иных правил упрощения, определять какие из отношений, типов ресурсов, их атрибутов ресурсов должны использоваться по умолчанию. Правила упрощения дают возможность достаточно просто формулировать сложные поисковые запросы, хотя и язык SL не предназначен для непосредственного использования, поскольку для этого имеются специальные визуальные средства, упрощающие создание поисковых запросов различных уровней сложности.

### 2.3. Поиск

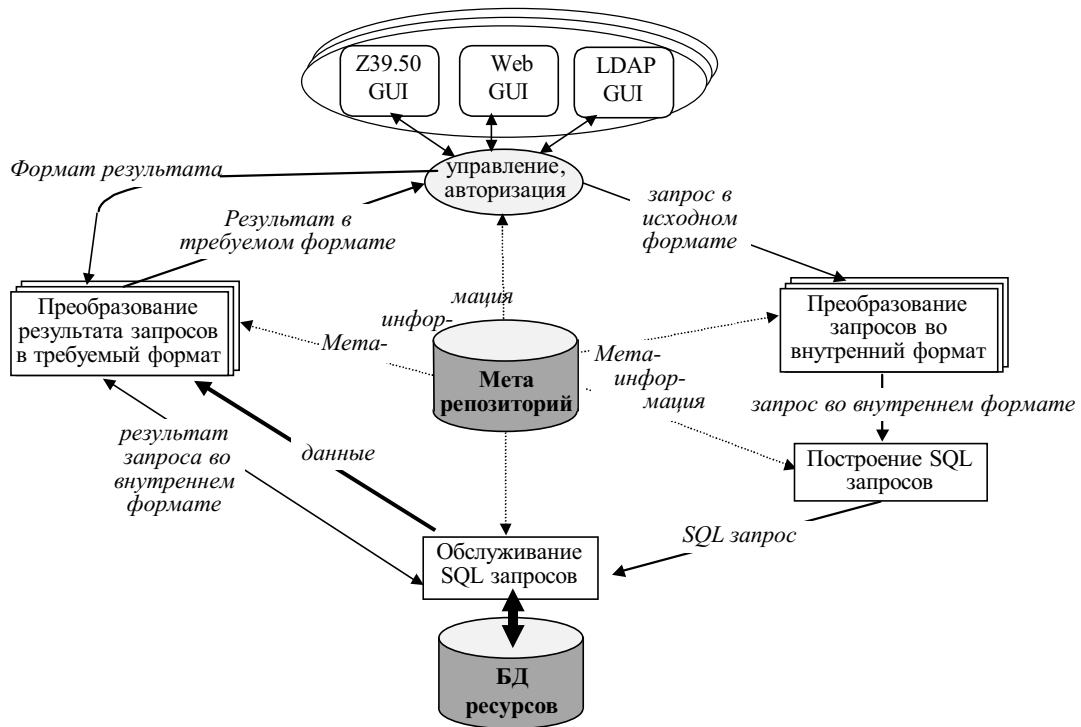


Рис. 4.

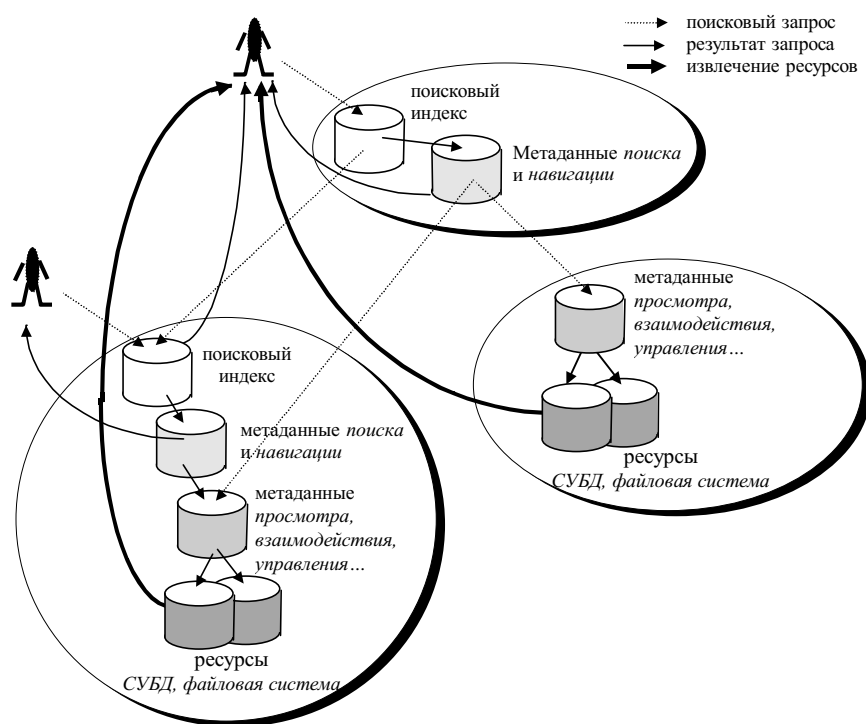
### Преобразование поисковых запросов

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, сопоставимые по мощности со всем множеством ресурсов системы, избавиться от сохранения промежуточных результатов. Преобразование исходных поисковых запросов (рис. 4) во внутренний формат и генерация SQL запросов производятся на основе метаинформации из метарепоzitория, описывающей предметную область и характеризующей функциональность поискового протокола.

Формирование результатов запроса также осуществляется на основе внутреннего представления. Это позволяет избежать накладных расходов 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.

### 2.4. Организация распределенного хранения ресурсов, их поиска и доступа к ним

Приведенная выше схема реализации ИСИР предназначена для реализации одного узла распределенной системы. В общем случае система может состоять из нескольких узлов (серверов). Система серверов, составляющих ИСИР, должна обеспечивать различные уровни работы с ресурсами (рис. 5), в зависимости от возможностей и политики организации, сопровождающей сервер.



**Рис. 5. Схема организации распределенных поиска и хранения**

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 ИСИР, основанные на открытых стандартах.

В 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. [7] и призванного идентифицировать информационные ресурсы вне зависимости от их местоположения в Интернет. URN является постоянным и уникальным идентификатором и 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 организации. 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 местоположениях 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 [8], DOI [9].

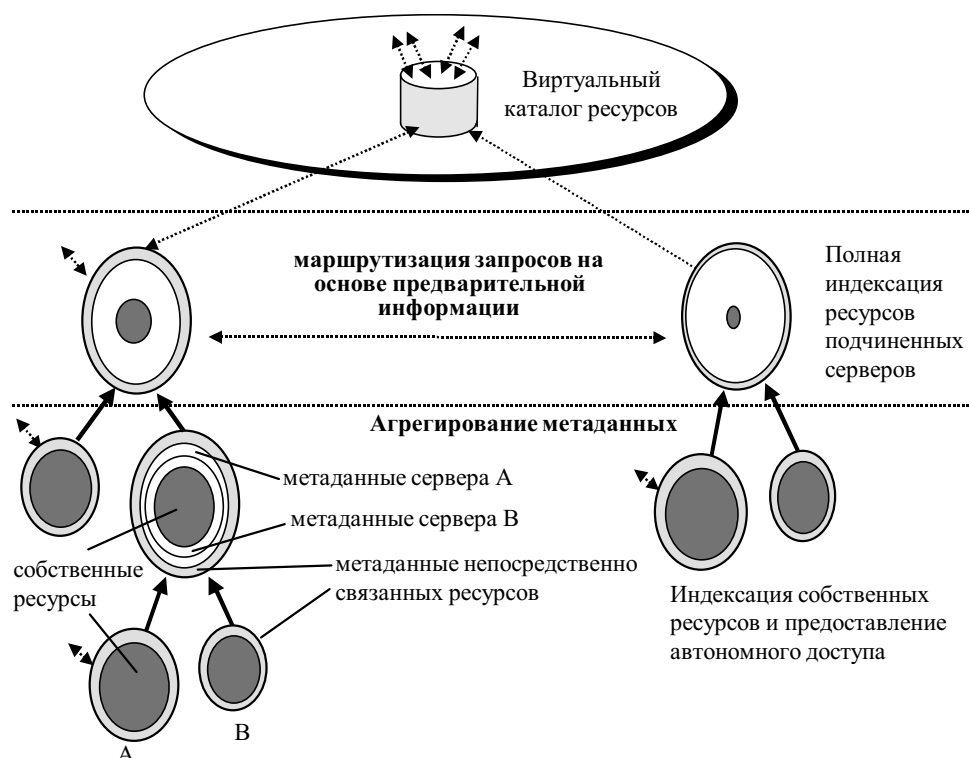


В ИСИР каждому серверу назначается собственный сегмент именования в зарегистрированном пространстве имен. В рамках этого сегмента организация отвечает за уникальность имен своих ресурсов. Вышестоящая организация может выделить часть своего сегмента своим подчиненным организациям.

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 изменениях). 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 этой персоны, а второй - копию поисковых метаданных публикации.
- Для организации эффективного поиска и обеспечения отказоустойчивости, особенно, при малопроизводительных 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 [10].
- 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) [11] позволяют оптимизировать прохождение распределенных запросов при такой организации взаимодействия.

Распределение данных и метаданных серверов, составляющих ИСИР, иллюстрируются рис. 6.



**Рис. 6. Распределение данных и метаданных**

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. ИСИР РАН

В настоящее время по изложенной схеме реализована 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 ЦБ. 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 (S) и MS Windows/NT, использующие РСУБД Oracle и Oracle Web-сервер, и MS Windows/NT, использующая РСУБД MS SQL Server и IIS. 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 (Java) метаданных формата Dublin Core, представленных в XML[12] синтаксисе методики RDF[13,14].

### Литература

- [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] The World Wide Web Consortium Home Page, <http://www.w3.org/>.
- [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] Sollins, K., and L. Masinter. "Functional Requirements for Uniform Resource Names", RFC 1737, MIT/LCS, Xerox Corporation, December 1994.
- [8] The Handle System Home Page, <http://www.handle.net/>.
- [9] International DOI Foundation web site; <http://www.doi.org>.
- [10] Carl Logoze, David Fielding. "Defining Collections in Distributed Digital Libraries", D-Lib Magazine, November 1998.
- [11] John Kirriemuir, Dan Brickley, Susan Welsh, Jon Knight, and Martin Hamilton. "Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach". D-Lib Magazine, January 1998; <http://www.dlib.org/dlib/january98/01kirriemuir.html>.
- [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] Dan Brickley, R.V. Guha and Andrew Layman, "Resource Description Framework, (RDF) Schemas", W3C Working Draft; <http://www.w3.org/TR/PR-rdf-schema>.

## Сведения об авторах.

*Бездушный Анатолий Николаевич*, кандидат физико-математических наук, старший научный сотрудник Вычислительного Центра РАН. Область основных интересов – распределенные информационные системы, цифровые библиотеки.

*Жижченко Алексей Борисович*, доктор физико-математических наук, Зам. Академика секретаря отделения математики РАН, Директор Центра научных телекоммуникаций и информационных технологий РАН. . Область основных интересов – распределенные информационные системы, телекоммуникационные сети, компьютерные, сети алгебраическая геометрия, информатика.

*Кулагин Михаил Викторович*, кандидат технических наук, заместитель директора Центра научных телекоммуникаций и информационных технологий РАН. . Область основных интересов – распределенные информационные системы, компьютерные сети, средства обработки данных.

*Серебряков Владимир Алексеевич*, доктор физико-математических наук, заведующий отделом Вычислительного Центра РАН. Область основных интересов – распределенные информационные системы, цифровые библиотеки.