

# Электронная библиотека научных информационных ресурсов КарНЦ РАН: состояние и перспективы развития

© Вдовицын В.Т., Сорокин А.Д., Луговая Н.Б.

Институт прикладных математических исследований КарНЦ РАН  
vdov@krc.karelia.ru, sorokin@krc.karelia.ru

## Аннотация

В статье представлены результаты работы по созданию инфраструктуры и содержания (контента) электронной библиотеки информационных ресурсов Карельского научного центра РАН – <http://dl.krc.karelia.ru>. Приводится обоснование предложенных технологических подходов и решений, а также рассматриваются возможные пути развития этой системы.

## 1 Введение

В настоящее время все более широкое распространение в сети Интернет получают электронные библиотеки (ЭБ), которые создаются в самых разнообразных областях человеческой деятельности и в конечном итоге должны стать распределенными хранилищами знаний. Много внимания уделяется исследованиям и разработкам в этом направлении и в России, в частности, с 1999 года ежегодно проводятся Всероссийские научные конференции RCDL, создан портал "Российские электронные библиотеки" <http://www.elbib.ru>, проекты создания электронных коллекций и библиотек поддерживаются РФФИ. Актуальной проблемой является создание ЭБ в научных учреждениях РАН, которые могут стать как средством интеграции разнородных электронных информационных ресурсов, так и местом для публикации и хранения новой научной информации. Тем самым ЭБ могут способствовать эффективному распространению научных результатов среди заинтересованных пользователей и обеспечить информационную поддержку совместной работы коллектива исследователей [1, 2, 6, 7].

В данной статье рассматриваются вопросы, связанные с разработкой и реализацией электронной библиотеки научных информационных ресурсов Карельского научного центра РАН на примере описания биологических объектов Карелии.

Труды 6<sup>ой</sup> Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" - RCDL2004, Пущино, Россия, 2004.

Приводится обоснование предложенных решений и рассматриваются возможные пути развития системы.

## 2 Инфраструктура ЭБ КарНЦ РАН

### 2.1 Основные принципы построения системы

Электронная библиотека Карельского научного центра Российской академии наук (ЭБ КарНЦ РАН) предназначена для формирования, хранения и многоцелевого использования коллекций электронных научных информационных ресурсов, создаваемых учеными центра на основе результатов многолетних исследований, и предоставления к ним доступа в удобном для пользователей виде через сети Интернет.

В основу разработки ЭБ КарНЦ РАН положены следующие основные особенности и принципы:

- ЭБ представляет собой систему, объединяющую в единое целое: коллекции электронных научных информационных ресурсов; сервисы (т.е. программное обеспечение для поддержки процессов публикации, сопровождения, хранения и поиска документов в коллекциях библиотеки); специалистов, обеспечивающих создание и сопровождение коллекций и сервисов, а также пользователей информационных ресурсов электронной библиотеки.
- Научная информация, на основе которой создаются коллекции, является разнородной (тексты, электронные карты-схемы, формулы, графические объекты и т. п.) и слабо структурированной.
- Специалисты-предметники должны, на наш взгляд, в максимальной степени самостоятельно осуществлять формирование, публикацию и сопровождение своих коллекций с учетом общепринятых в их научной среде стандартов и в рамках определенного для них регламента работы, а также должны

обеспечивать качество и достоверность информационного содержания своих документов.

- Сервисы ЭБ должны включать удобные для специалистов-предметников и эффективные с точки зрения реализации программные средства автоматизации процессов формирования, публикации и сопровождения своих коллекций с учетом разграничения их полномочий и защиты информационных ресурсов от несанкционированного доступа.
- При разработке сервисов ЭБ и создании научных коллекций информационных ресурсов необходимо использовать общепринятые стандарты на интерфейсы и метаданные, а также эффективные информационные технологии.
- Пользователи ЭБ должны иметь удобные средства для доступа к нужной информации по запросам. При этом в ЭБ КарНЦ РАН должна быть задействована функция "точки входа" в единое информационное пространство центра.
- ЭБ КарНЦ РАН должна разрабатываться и развиваться как самостоятельная информационная система в соответствии со своими целями и задачами и в то же время ее разработка и развитие должны быть увязаны с созданием других академических информационных систем, в первую очередь, с ИСИР РАН [2].

## 2.2 Описание основных сервисов

Исходя из сформулированных принципов и особенностей построения ЭБ КарНЦ РАН, мы выделили ряд следующих категорий пользователей:

- администратор предметной коллекции научных информационных ресурсов;
- авторы документов предметной коллекции;
- эксперты предметной коллекции;
- пользователи информационных ресурсов ЭБ.

Первые три категории пользователей регистрируются на головном сайте ЭБ КарНЦ РАН и получают соответствующие права работы с предметной коллекцией. Процесс формирования, публикации и сопровождения коллекции информационных ресурсов в нашем случае происходит по следующей схеме. Во-первых, администратор и авторы документов предметной коллекции разрабатывают паспорта описания объектов, в которых структурируется разнородная научная информация об исследуемых объектах. Эти паспорта обсуждаются и согласовываются с экспертами предметной коллекции. Во-вторых, на основе разработанных и согласованных паспортов "вручную" формируется соответствующее DTD-определение структуры XML-до-

кументов новой коллекции. В-третьих, администратор коллекции совместно с авторами документов и экспертами коллекции описывает общие свойства предметной коллекции на основе атрибутов стандарта Дублинского ядра (DC, Dublin Core, <http://purl.org/dc/documents>). После этого происходит ввод (корректировка) документов в коллекцию и обсуждение их информационного содержания с экспертами на форуме ЭБ.

Привлечение специалистов-предметников к непосредственному участию в процессе создания и сопровождения научной коллекции в ЭБ КарНЦ РАН способствует, на наш взгляд, улучшению качества и достоверности информационного содержания документов и повышает их заинтересованность в конечном результате работы. По существу, каждый документ коллекции следует рассматривать как электронную публикацию, в которой в сжатой форме представлена обобщающая научная информация по результатам многолетних исследований описываемого объекта коллекции. Использование XML-платформы при разработке ЭБ КарНЦ РАН позволяет нам применить всю мощь этой перспективной технологии, преимущества которой отмечены, в частности, в статьях М. Р. Коголовского [6, 7]. Представление коллекций электронных информационных ресурсов в XML-формате позволяет осуществить автоматическую верификацию целостности документов, а также разработать интерфейсные формы для заполнения полей документов самими специалистами-предметниками. Применение стандарта DC для описания общих свойств каждой предметной коллекции используется нами для создания каталога коллекций, предназначенного для организации поиска коллекции по запросу пользователя.

Таким образом, основные сервисы ЭБ КарНЦ РАН связаны с поддержкой процессов публикации и сопровождения коллекций научных информационных ресурсов, а также с поиском документов в коллекциях библиотеки по запросам пользователей [4]. Поддержка работы специалистов-предметников по вводу и корректировке документов в коллекции ЭБ осуществляется за счет разработанного дружественного интерфейса, автоматической верификации вводимого документа заданному DTD-определению коллекции и автоматического формирования действительного XML-документа, а также организация процесса обсуждения его на форуме [3]. Процедура поиска данных в ЭБ КарНЦ РАН включает два этапа - поиск нужной коллекции среди других коллекций электронной библиотеки и поиск документа в выбранной коллекции по запросу пользователя. При этом поиск коллекции осуществляется как при помощи рубрикатора ГРНТИ, так и по базе метаданных, описывающей общие свойства коллекций ЭБ. Запросы на поиск документов в выбранной коллекции формируются пользователем с помощью специально разработанных интерфейсных форм путем указания значений атрибутов искомого документа. При этом список полей,

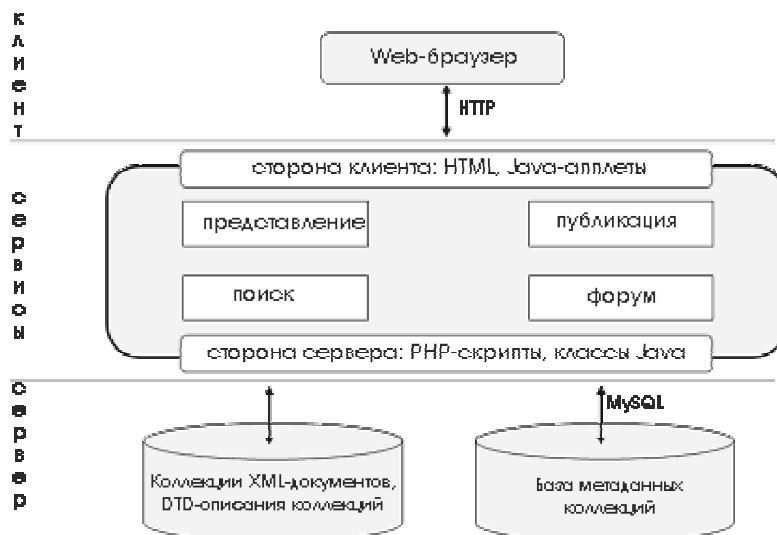


Рис. 1: Общая архитектура электронной библиотеки КарНЦ РАН

по которым возможен отбор документов, специфичен для каждой коллекции [5].

### 2.3 Особенности реализации

Общая архитектура ЭБ КарНЦ РАН представлена на рис. 1. Для реализации сервисов ЭБ использованы традиционные Интернет-технологии, XML, СУБД MySQL, язык PHP и технология CGI-сценариев, а также язык Java и Java API for XML. Язык XML используется для хранения документов коллекций.

Для представления документов в браузере разработаны специальные PHP-скрипты и применяется язык преобразования XSLT. С помощью СУБД MySQL на основе атрибутов стандарта DC создана база данных, описывающая общие свойства коллекций ЭБ.

Поиск нужной коллекции осуществляется по запросу пользователя с помощью разработанных с применением CGI-технологии интерфейсных форм. Возможны два варианта поиска:

- Поиск по рубриктору. В этом случае поиск коллекции может быть осуществлен как по ключевому слову, так и по каталогу отраслей науки, составленному на основании рубрикатора ГРНТИ (рис. 2).
- Поиск по базе метаданных. В этом случае поиск может быть осуществлен путем указания значений некоторых свойств коллекции (рис. 3).

The screenshot shows the 'Поиск по рубриктору' interface. At the top, it says 'Введите ключевое слово' (Enter the keyword) with an input field and a 'найти' (find) button. Below that, it says 'или выберите в каталоге' (or choose in the catalog). A tree view is shown with the following structure:

- ▼ Биология
  - ▶ Микробиология
  - ▼ Ботаника
    - Общие вопросы
    - Споровые растения
    - Семенные растения
    - Растительность. Фитоценология
  - ▶ Физиология растений
  - ▶ Экология

Рис. 2: Поиск коллекции по рубриктору

The screenshot shows the 'Поиск по базе метаданных' interface. It contains several input fields and buttons:

- Название коллекции (Collection name)
- Сотрудник или подразделение КарНЦ РАН, ответственные за подготовку коллекции (Employee or department of the Institute of Botany of the Russian Academy of Sciences, responsible for the collection preparation)
- Ключевые слова или фразы описывающие содержание коллекции (Keywords or phrases describing the collection content)
- Дата создания или открытия доступа к коллекции (гггг/мм/дд) (Creation or access date in YYYY/MM/DD format)
- Язык представления коллекции (Collection presentation language)
- Buttons: 'поиск' (search) and 'очистить' (clear)

Рис. 3: Поиск коллекции по базе метаданных

В результате поиска выводится список найденных коллекций с описанием их общих свойств.

В виду того, что коллекция может содержать не одну сотню документов, для удобства пользователю предоставляется возможность поиска документов в коллекции с помощью интерфей-



Рис. 4: Просмотр документа из коллекции “Млекопитающие Карелии”

ной формы, которая генерируется на основе отобранных поисковых полей для каждой коллекции (рис. 5).

Поиск документов коллекции  
“Млекопитающие Карелии”

Название на русском языке

Название на латинском языке

Отдел на русском языке

Класс на русском языке

Семейство на русском языке

Род на русском языке

Вид на русском языке

Рис. 5: Поиск документов коллекции

В результате поиска на экран выводится список названий найденных документов со ссылками

на их полное описание (рис. 4).

Поиск искомой коллекции в ЭБ по базе данных MySQL реализован на языке PHP, а по XML-документам коллекции - с использованием модели DOM.

Процесс публикации документов осуществляется сотрудником, ответственным за коллекцию. После регистрации, этот сотрудник получает доступ к сервису публикации, с помощью которого он может работать с соответствующей коллекцией - корректировать метаописание, добавлять новые документы или изменять существующие. Авторизация пользователей организована путем отслеживания сессий PHP. При этом у администратора ЭБ есть возможность отслеживать статистику посещений сервиса публикации. Публикация и сопровождение коллекции реализованы с использованием технологий PHP, Java, MySQL и XML. Изменение метаописания коллекции реализовано на PHP в виде запросов к базе данных MySQL (рис. 6). Добавление и редактирование описания документов реализовано в виде апплета Java на стороне клиента (рис. 7) и Java-приложения на стороне сервера. При этом производится разбор DTD-документа коллекции и на его основании строится дерево документа. Загрузка картинок происходит посредством скриптов PHP (рис. 8).

Публикация документов коллекции

Изменение коллекции отв. Иванов П.С.

Название	Ответственный	Ключевые слова	Дата	Язык
<input checked="" type="radio"/> Млекопитающие Карелии	Лаборатория зоологии, Институт биологии КарНЦ РАН	распространение, численность, пространственное распределение, особенности экологии, охрана, управление популяциями, роль в биоценозах	2004-05-24	русский

изменить метаописание     удалить метаописание  
 добавить документ     изменить документ

Меню ответственного за коллекцию

Добавить коллекцию  
 Редактировать коллекцию

Рис. 6: Изменение коллекции

Редактирование документа

Элемент Вид Завершить работу

- описание вида
  - название на русском языке
  - название на латинском языке
  - название на английском языке
  - название на финском языке
  - название на шведском языке
  - название на норвежском языке
  - название на немецком языке
- классификация
  - класс на русском языке
  - класс на латинском языке
  - отряд на русском языке
  - отряд на латинском языке
  - семейство на русском языке
  - семейство на латинском языке
  - подсемейство на русском языке
  - подсемейство на латинском языке
  - род на русском языке
  - род на латинском языке
  - вид на русском языке
  - вид на латинском языке
  - подвид на русском языке
  - подвид на латинском языке
- морфология
  - внешний вид

Волк

Рис. 7: Редактирование документа

### 3 Направления развития системы

Возможные направления развития ЭБ КарНЦ РАН связаны, на наш взгляд, с расширением информационного содержания (контента) и совершенствованием программного обеспечения для поддержки основных сервисов системы. Расширение контента ЭБ предполагает:

- создание новых научных коллекций информационных ресурсов по разработанной технологии и пополнение существующих коллекций новыми документами;
- описание в стандарте DC всех ранее созданных в центре научных электронных инфор-

Canis lupus L., 1758 - Волк

Рисунки (1)  
 1. Волк (Canis lupus L., 1758)  
 ширина: 300 высота: 250  
 Файлы формата JPEG или GIF (с прозрачным фонком), с шириной не больше 400 пикселей и объемом не более 70 КБ




Рис. 1. Волк (Canis lupus L., 1758)

удалить рисунок

новая картинка:   
 подписанная подпись:   
 ширина:  высота:

Зубные формулы (1)  
 1.   
 ширина: 100 высота: 34  

$$\begin{matrix} 1 & 3 & 1 & 4 & 2 \\ 1 & - & C & - & P & - & M & - & 2 \\ 3 & 1 & 4 & 3 \end{matrix}$$
 Файлы формата GIF с прозрачным фонком и с шириной не больше 200 пикселей и объемом не более 30 КБ

удалить рисунок

новая картинка:   
 подписанная подпись:   
 ширина:  высота:

Рис. 8: Загрузка картинки

мационных ресурсов и пополнение базы метаданных ЭБ. В этом случае поисковые средства ЭБ КарНЦ РАН смогут обеспечить доступ пользователей по запросам к электронным информационным ресурсам центра;

- создание коллекции электронных публикаций ученых центра и организация доступа к ним по запросам пользователей;
- формирование и подключение к поисковой машине ЭБ списков ссылок в Интернете на коллекции научных электронных информационных ресурсов, созданных в других организациях и соответствующих тематике исследований центра.

Развитие программного обеспечения для поддержки основных сервисов ЭБ предполагает, в первую очередь, решение следующих основных задач. Во-первых, необходимо разработать и реализовать эффективную модель хранения и поиска XML-документов в ЭБ КарНЦ РАН. Во-вторых, с учетом пожеланий специалистов-предметников и возможностей современных информационных технологий можно попытаться повысить степень автоматизации и надежности поддержки процессов публикации и сопровождения документов в коллекциях ЭБ КарНЦ РАН. В-третьих, целесообразно исследовать и оптимизировать характеристики функционирования программного обеспечения системы в целом.

#### 4 Заключение

В настоящее время по разработанной технологии в ЭБ КарНЦ РАН - <http://dl.krc.karelia.ru> созданы и пополняются новыми документами три коллекции научных информационных ресурсов: "Аффилофороидные грибы Карелии" (свыше 150 документов), "Виртуальная флора Карелии" (свыше 75 документов) и "Млекопитающие Карелии" (свыше 20 документов). Основой создания электронных коллекций явился большой объем научной информации, собранной сотрудниками институтов биологии и леса Карельского научного центра РАН. Для видов растений в документы коллекции включены данные по систематике, морфологии, экологии, распространению, хозяйственному значению и охране видов. Документы коллекции по млекопитающим содержат информацию по систематике, морфологии (внешний вид, размеры тела и черепа, качество меха), биологии (распространение, динамика численности, размножение, рост, развитие, поведение и т.д.), а также отмечена роль животных в биоценозах и их хозяйственное значение. На основе многолетнего мониторинга представлены карты-схемы распределения растений и животных на территории Карелии. Описания биологических объектов сопровождаются их изображениями. Также прилагается список литературы по данной тематике.

Авторы выражают свою благодарность Бедореву А. А., Гудковой И. В., Чуйко Ю. В. за их вклад в разработку и сопровождение программных сервисов ЭБ КарНЦ РАН, а также специалистам институтов леса и биологии центра - Крутову В. И., Руоколайнен А. В., Кузнецову О. Л., Данилову П. И., Федорову Ф. В., Знаменскому С. Р. за подготовку и ввод документов в коллекции и апробацию разработанной технологии.

Данная работа выполняется при финансовой поддержке РФФИ (грант N 02-07-90204)

#### Список литературы

[1] Арнаутов С. А. Роль и место виртуальных цифровых библиотек в Интернете. // Сб. докл. Третьей Всеросс. конф. "Электронные

библиотеки: перспективные методы и технологии, электронные коллекции". RCDL'2001, Петрозаводск, 11-13 сентября 2001 г. с. 20-25.

- [2] Бездушный А.Н., Жижченко А.Б., Кулагин М.В., Серебряков В.А. Интегрированная система информационных ресурсов РАН и технология разработки электронных библиотек. // Программирование. N 4, 2000 г., с. 177-185.
- [3] В.Т. Вдовицын, А.Д. Сорокин Технология публикации и сопровождения документов в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН // Труды пятой Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург, 29-31 октября 2003 г. с.103-105.
- [4] Вдовицын В.Т., Сорокин А.Д. Вопросы построения электронной библиотеки Карельского научного центра РАН. // Труды Четвертой Всеросс. конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". RCDL'2002. ( Дубна, 15-17 октября 2002 г. ) в 2-х т.-Дубна, ОИЯИ, -Т. 1, с. 304-308.
- [5] Вдовицын В.Т., Сорокин А.Д., Луговая Н.Б., Беляева Н.А. Поиск данных в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН. //Материалы 8 Межд. конф. по электронным публикациям. "Ei-Pub 2003", 8-10 октября 2003 г., г. Новосибирск.
- [6] Коголовский М.Р. Систематика коллекций информационных ресурсов в электронных библиотеках. // Программирование. N 3, 2000 г. с. 31-52.
- [7] Коголовский М.Р. Стандарты XML и электронные библиотеки. Электронные библиотеки - 2003 - Том 6 - Выпуск 2.

#### Digital Library of Scientific Information of the Karelian Research Centre of the Russian Academy of Science: status and perspectives

Vladimir V. Vdovitsyn, Anatoly D. Sorokin,  
Natalia B. Lugovaya

The paper presents the results of the ongoing project for creating the Karelian Research Centre digital library infrastructure and contents-<http://dl.krc.karelia.ru>. The theoretical grounds of the proposed approaches and technical solutions are given. Potential ways for further improvement of the digital library are discussed.