

# Технология публикации и сопровождения документов в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН

© Вдовицын В.Т., Сорокин А.Д.

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

## Аннотация

В статье представлена разработанная и реализованная технология публикации и сопровождения документов в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН – <http://dl.krc.karelia.ru>. Система предназначена для упрощения работы специалистов-предметников по вводу и корректировке документов в коллекции электронной библиотеки, а также повышения качества и достоверности информационного содержания этих документов.

## 1 Введение

В настоящее время одним из эффективных средств информационной поддержки научных исследований и образования являются электронные библиотеки (ЭБ), содержащие коллекции научных информационных ресурсов в определенных областях знаний и программные средства, обеспечивающие публикацию, сопровождение и поиск документов в коллекциях библиотеки, а также их доступность и сохранность [3]. Перспективным направлением является создание электронных библиотек в Российской академии наук, которые могут стать хранилищами всей информации, создаваемой в учреждениях РАН [1]. В связи с этим актуальной является задача построения ЭБ Карельского научного центра РАН, которая должна, на наш взгляд, стать основой при создании единой информационной системы центра и в перспективе может быть интегрирована в Единую Информационную Систему РАН.

## 2 Электронная библиотека КарНЦ РАН

Электронная библиотека КарНЦ РАН предназначена для формирования, хранения и многоцелевого использования различного рода коллекций

научных электронных информационных ресурсов, создаваемых учеными центра, и предоставления к ним доступа в удобном для пользователей виде через сети Интернет. В данной статье представлены результаты, которые являются продолжением работы над проектом создания ЭБ КарНЦ РАН [2]. Близкими по сути аналогами такого рода информационных систем являются, например, системы ИСИР РАН и ИРИС СО РАН [4].

В основу разработки ЭБ КарНЦ РАН положены следующие принципы:

- специалисты-предметники должны самостоятельно осуществлять формирование и сопровождение своих коллекций электронных научных информационных ресурсов с учетом общепринятых в научной среде стандартов и в рамках определенного для них регламента работы с документами;
- программное обеспечение электронной библиотеки должно включать удобные для специалистов-предметников средства автоматизации процессов формирования и сопровождения электронных информационных ресурсов.

К настоящему времени спроектирована и реализована первая версия головного сайта ЭБ центра – <http://dl.krc.karelia.ru>, на котором размещено разработанное программное обеспечение для поддержки процессов публикации, сопровождения и поиска документов в коллекциях библиотеки. Созданы и пополняются документами три коллекции научных электронных информационных ресурсов – “Аффилофороидные грибы Карелии”, “Виртуальная флора Карелии” и “Млекопитающие Карелии”. При этом структуры документов каждой коллекции разработаны на основе представленных специалистами-предметниками паспортов описаний биологических объектов и реализованы с помощью XML. Для описания общих свойств каждой коллекции используется стандарт Дублинского ядра.

Поиск данных в ЭБ КарНЦ РАН включает два основных этапа – поиск искомой коллекции

среди других коллекций электронной библиотеки и поиск документа(ов) в найденной коллекции по запросу пользователя. Поиск коллекции может быть осуществлен как при помощи рубрикатора ГРНТИ, так и с помощью базы метаданных, описывающих общие свойства каждой коллекции на основе атрибутов стандарта DC. При этом предполагается задействовать в поисковом сервисе только те разделы рубрикатора ГРНТИ, по тематике которых ведутся исследования в институтах центра.

Запросы на поиск документов в выбранной коллекции формируются пользователем с помощью разработанных интерфейсных форм путем задания значений параметров документа, специфичных для каждой коллекции. При этом используется техника поиска, основанная не на методе сплошного полнотекстового сканирования, а на использовании структурных и семантических свойств рассматриваемых документов.

В перспективе предполагается, что наряду с расширением состава и количества документов коллекций, создаваемых специалистами по данной технологии, с помощью атрибутов стандарта DC будут описаны и включены в базу метаданных ЭБ КарНЦ РАН все созданные в институтах центра электронные научные информационные ресурсы со ссылками на другие аналогичные по тематике российские и зарубежные коллекции. Тем самым в ЭБ КарНЦ РАН будет задействована функция “точки входа” в единое электронное информационное пространство центра.

### 3 Технология публикации и сопровождения коллекций

Главная цель разработки и реализации предлагаемой на рассмотрение технологии публикации и сопровождения документов в коллекциях электронной библиотеки КарНЦ РАН состоит в том, чтобы:

- облегчить специалистам-предметникам работу по вводу и корректировке документов в коллекциях электронной библиотеки;
- повысить качество и достоверность информационного содержания документов коллекций.

Упрощение работы специалистов-предметников по вводу и корректировке документов осуществляется за счет разработки дружественного интерфейса, автоматической проверки вводимого документа заданному DTD-определению и автоматического формирования соответствующего корректного XML-документа коллекции. Тем самым специалисты-предметники при формировании и сопровождении документов коллекции могут сосредоточиться в первую очередь на содержательной части документа(ов). Им нет необходимости детально знакомиться со спецификациями стандарта XML и самостоятельно осуществлять проверку корректности создаваемого или изменяемого XML-документа.

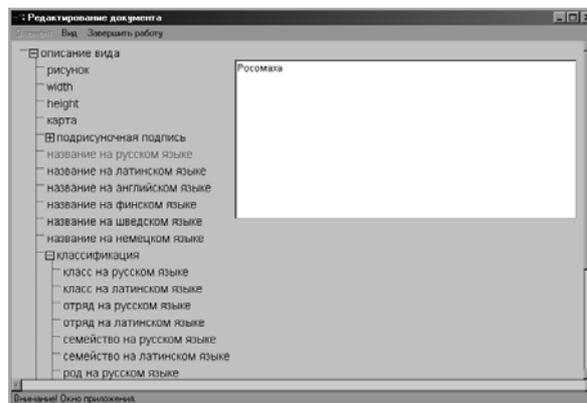


Рис. 1: Редактирование документа из коллекции “Млекопитающие Карелии”

На Рис. 1 представлен интерфейс окна апплета, порожденного DTD-определением документов коллекции “Млекопитающие Карелии”.

Повышение качества и достоверности информационного содержания документов коллекции осуществляется путем организации и проведения на форуме ЭБ КарНЦ РАН обсуждения экспертами каждого вводимого в коллекцию документа. Это позволит, на наш взгляд, достаточно оперативно осуществлять рецензирование публикуемых в Интернет электронных материалов.

В рамках разработанной технологии в ЭБ выделены следующие категории пользователей:

- **АДМИНИСТРАТОР КОЛЛЕКЦИИ** – осуществляет совместно со специалистами-предметниками создание метаописания коллекции, а также просмотр, ввод, корректировку и удаление любого документа коллекции;
- **АВТОР ДОКУМЕНТА КОЛЛЕКЦИИ** – осуществляет ввод и просмотр “своего” документа коллекции, а также корректировку и удаление “своего” документа с санкции администратора коллекции;
- **ЭКСПЕРТЫ КОЛЛЕКЦИИ** – осуществляют просмотр и обсуждение на форуме информационного содержания документов коллекции;
- **ПОЛЬЗОВАТЕЛИ КОЛЛЕКЦИИ** – осуществляют поиск и просмотр документов в коллекциях электронной библиотеки.

Таким образом, технологию публикации и сопровождения документов в ЭБ КарНЦ РАН в самом общем виде можно представить следующим образом. Во-первых, на основе согласованного среди специалистов-предметников паспорта (описания) объекта формируется соответствующее DTD-определение документов новой коллекции, которое помещается в базу метаданных электронной библиотеки. Во-вторых, администратор коллекции совместно с авторами документов описывает общие свойства коллекции с использованием атрибутов стандарта Дублинского ядра. В-

третьих, администратор коллекции и авторы документов регистрируются в ЭБ и получают соответствующие права работы с данной коллекцией. Все эти действия выполняются во взаимодействии с системным администратором, сопровождающим программное обеспечение ЭБ.

Пополнение коллекции документами осуществляется по следующей схеме:

1. автор документа с помощью разработанной интерфейсной формы вводит очередной документ и отправляет его в коллекцию. При этом осуществляется проверка вводимого документа на соответствие DTD-определению и автоматически формируется корректный XML-документ;
2. данный документ временно помещается на форум ЭБ КарНЦ РАН. При этом автоматически посылаются e-mail-сообщения о появлении на форуме коллекции нового документа администратору и экспертам коллекции;
3. в течение определенного времени происходит обсуждение содержания документа на форуме (т.е. эксперты пишут отзывы и дают свои замечания, которые отправляются автору документа и администратору коллекции по электронной почте). Окончательное решение о внесении документа в коллекцию принимает администратор с учетом проведенного обсуждения.

#### 4 Заключение

Разработанная технология публикации и сопровождения документов в коллекциях ЭБ КарНЦ РАН реализована с использованием Java, Java API for XML, а также набора CGI-скриптов, обеспечивающих интерфейс с СУБД MySQL. Авторы статьи выражают свою благодарность Луговой Н.Б., Беляевой Н.А., Гудковой И.В. и Чуйко Ю.В. за их труд по реализации программного обеспечения ЭБ КарНЦ РАН.

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] 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.

- [3] Коголовский М. Р. Систематика коллекций информационных ресурсов в электронных библиотеках. // Программирование. N 3, 2000 г. с. 31–52.

- [4] Шокин Ю. И., Федотов А. М., Леонова Ю. В. Принципы динамического формирования документов в информационных системах, на примере интегрированной распределенной информационной системы (ИРИС) СО 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. 2, с. 159–169.

### The technology of electronic documents publication and maintainance in the collections of informational resources of the digital library of KRC RAS

Vdovitsyn V.T., Sorokin A.D.

The article concerns the description of the technology of electronic documents publication and maintainance in the collections of scientific informational resources of the digital library of Karelian Research Centre – <http://dl.krc.karelia.ru>. The system was made to simplify the work of the specialists in particular spheres of science who want to input and edit data in the collection of digital library. The other task was to raise the quality and authenticity of the documents content.