

УДК 681.3

ЭЛЕКТРОННЫЕ НАУЧНЫЕ ИНФОРМАЦИОННЫЕ РЕСУРСЫ КАРЕЛЬСКОГО НАУЧНОГО ЦЕНТРА РАН: СОСТОЯНИЕ, ПРОБЛЕМЫ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ

В. Т. ВДОВИЦЫН, А. Д. СОРОКИН

Институт прикладных математических исследований Карельского научного центра РАН

В статье представлен обзор основных направлений исследований и разработок по созданию и эффективному использованию электронных научных информационных ресурсов для поддержки фундаментальных исследований, образования и инновационной деятельности с применением Интернет-технологий. Рассматриваются вопросы создания Единой информационной системы КарНЦ РАН как составной части Единого Научного Информационного Пространства Российской академии наук. Особое внимание уделено построению электронной библиотеки научных информационных ресурсов КарНЦ РАН.

V. T. VDOVITSYN, A. D. SOROKIN. DIGITAL SCIENTIFIC INFORMATION RESOURCES OF KARELIAN RESEARCH CENTRE, RUSSIAN ACADEMY OF SCIENCE: STATE-OF-THE-ART, CHALLENGES AND POTENTIAL APPLICATIONS

The paper is a review of the principal research areas and developments related to generation and efficient use of digital scientific information resources meant to provide support to basic research, education and innovative activities using Internet technologies. Problems of establishing the Shared information system of the Karelian Research Centre of RAS as a component part of the Shared Scientific Information Space of the Russian Academy of Science are considered. Special focus is on building the digital library of scientific information resources for Karelian Research Centre of RAS.

Ключевые слова: научные информационные системы, электронные библиотеки и электронные коллекции, Интернет-технологии, Web-сайт, база данных.

Информационные системы для поддержки научных исследований

НАУЧНЫЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Научные информационные системы (НИС, Current Research Information Systems – CRIS) можно определить как информационные системы для доступа к научной информации. Этот класс систем достаточно подробно описан (Кулагин, Лопатенко, 2001; Лопатенко. Научные...; CERIF-2000). Здесь приводится краткое изложение некоторых основных сведений об этих системах.

Основными категориями пользователей НИС являются (Кулагин, Лопатенко, 2001):

– УЧЕНЫЕ (доступ к актуальной научной информации, поиск партнеров, финансирования, оборудования и вычислительных мощностей);

– ПРЕПОДАВАТЕЛИ И СТУДЕНТЫ (знакомство с достижениями науки и техники и использование результатов научных исследований в образовательном процессе);

– ИНВЕСТОРЫ И ПРОМЫШЛЕННОСТЬ (поиск перспективных научных разработок и новых технологий, а также экспертов, способных оценить новые научные результаты и технологии и способствовать их внедрению в производство);

– РУКОВОДИТЕЛИ ГОСУДАРСТВЕННЫХ ОРГАНОВ УПРАВЛЕНИЯ (управление наукой,

определение роли научных учреждений и ученых в научном процессе);

– ГРАЖДАНЕ (знакомство с достижениями науки и техники).

С учетом информационных потребностей перечисленных категорий пользователей выделяют следующие основные виды информационных ресурсов, с которыми работают НИС: информация об организациях, проектах и отчетах о проделанной работе; персональная информация; публикации; описание научных результатов, патентов, технологий и уникального научного оборудования; информация о научных фондах, электронных библиотеках, базах данных и вычислительных ресурсах, а также различная нормативно-справочная и библиографическая информация и другие документы, относящиеся к научной и научно-организационной деятельности учреждения.

Современные НИС должны удовлетворять следующим основным требованиям (Кулагин, Лопатенко, 2001; Лопатенко. Научные...). Во-первых, в этих системах необходимо организовать сбор соответствующей информации и перевод ее в электронную форму. При этом выделяют следующие процедуры ввода данных: интерактивный ввод данных пользователями; сбор данных в сети посредством специализированных сетевых «пауков» (мигрирующих по сети программ-агентов); путем обмена данными с другими НИС; интеграцией НИС с другими информационными системами. Во-вторых, обеспечение релевантности документов. Эта проблема обостряется при автоматическом сборе информации в сети Internet сетевыми «пауками». Релевантность документов обеспечивается точным указанием автоматическим поисковым средствам пространства поиска информации и критериев качества собранной информации; разделением всех информационных ресурсов на собранные экспертами и «пауком» с указанием степени достоверности. В-третьих, обеспечение актуальности, полноты и достоверности документов. При этом проблема достоверности информации решается: для интерактивного ввода – ограничением ввода только аутентифицированными пользователями; для автоматического сбора – ограничением области действия собирающего информацию «паука»; при обмене информацией с другими НИС – путем установления фильтров на импортируемые ресурсы. В-четвертых, наличие служб обслуживания запросов пользователя, включающих как поддержку поиска по атрибутам, так и обеспечение полнотекстового поиска и просмотра информационных ресурсов по категориям. Кроме основных к НИС предъявляются и дополнительные требования, связанные с поддержкой распределенной архитектуры, различных уровней абстракции хранимой информации, возможности архивирования информации. Некоторые специалисты отмечают важность свойства НИС выступать в

качестве научных порталов, например, система CORDIS (CERIF-2000), научный портал «Российская Научная Сеть» (<http://www.nature.ru/>).

Среди проблем создания и использования НИС наиболее существенными являются проблемы сбора необходимой информации, а также привлечение пользователей к работе с системой. Возможные пути решения – оплата информации, интеграция НИС с финансирующими научные исследования организациями (научными фондами), организация сетей обмена информацией с другими подобными системами на взаимовыгодной основе.

В качестве примеров российских информационных систем, которые можно отнести к классу НИС, следует отметить:

– информационный Web-портал РАН (<http://www.ras.ru/>) – обеспечивает доступ к информации об институтах, ученых, проектах и публикациях с организацией просмотра и поиска необходимой информации по запросам;

– ИС СО РАН (<http://www-sbras.nsc.ru/>) – содержит информацию о научной деятельности отделения, описание важнейших разработок, материалы научных конференций и т. п.;

– сервер ФГУ ГНИИ ИТТ «Информика» (<http://www.informika.ru/>) – содержит ряд баз данных (БД) по организациям, ученым, научным конференциям и т. п.

ЕДИНАЯ СРЕДА РАСПРЕДЕЛЕННЫХ РЕСУРСОВ

В настоящее время во многих странах мира активно разрабатывается концепция создания единой среды распределенных ресурсов (ЕСР, GRID), обеспечивающей глобальную интеграцию информационных и вычислительных ресурсов на основе новых компьютерных технологий с целью информационного обеспечения процессов принятия решений (более подробную информацию об этих системах можно найти, например, в работах А. В. Жучкова, С. А. Арнаутова (2001), В. А. Ильина, В. В. Корежкова (2002)). Выделяют два основных направления в развитии GRID-систем: вычислительный (computational) GRID и информационный (data intensive) GRID. При этом проекты создания вычислительных GRID-систем направлены на создание распределенных компьютерных систем, обеспечивающих максимальную скорость вычислений за счет глобального распараллеливания вычислений (например, проект создания распределенного суперкомпьютера TeraGRID, <http://www.teragrid.org>). Проекты информационных DataGRID-систем обеспечивают хранение, обработку и анализ огромных массивов данных (сотни терабайт) с одновременным доступом к информации тысяч пользователей по сети Internet. В качестве примера следует отметить масштабный международный проект EU Data GRID (<http://www.eu-datagrid.org>), направленный на создание глобальной инфраструктуры нового поколения для хранения и обработки

огромных массивов информации в области физики высоких энергий, биологии и систем наблюдения за Землей. В качестве основы для разработки промежуточного программного обеспечения (middleware) в проекте EU DataGRID выбран набор инструментальных средств Globus (<http://www.globus.org>). Таким образом, ECP (GRID)-систему можно представить в виде огромного гипотетического метакomпьютера, к которому могут обратиться пользователи по сети Internet по единому интерфейсу для получения необходимых для принятия решений информационных и вычислительных ресурсов.

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

Электронная библиотека (ЭБ, Digital Library) – распределенная информационная система, обеспечивающая формирование, хранение и эффективное использование различного рода коллекций электронных информационных ресурсов и предоставляющая доступ к ним в удобном для пользователя виде через глобальные сети передачи данных – Internet. Коллекция электронных (цифровых) информационных ресурсов (Information Resources Collection) – систематизированная совокупность информационных ресурсов, объединенных по какому-либо критерию, например, по общности содержания, источников, назначения, авторства, круга пользователей, владельцев, способов доступа и т. п. (Когаловский, 2000, 2002; Армс, 2001).

Исследования и разработки в области создания и использования электронных библиотек активно ведутся в различных странах мира. Регулярно проводятся ежегодные научные конференции, например, европейская конференция ECDL, а также объединенная конференция ACM/IEEE по электронным библиотекам – JCDL. В России проведено семь научных конференций по электронным библиотекам. Следует отметить, что в ряде публикаций отмечается перспективность применения этой технологии при создании информационных систем в научных учреждениях.

В работе С. А. Арнаутова (2001) рассматриваются перспективы создания ЭБ в Российской академии наук, которые потенциально могут стать универсальными хранилищами всей информации, создаваемой учреждениями РАН. При этом ключевыми функциями таких библиотек являются обеспечение интеграции разнородных информационных ресурсов и публикация всех видов научной информации, а также обеспечение ее доступности и сохранности.

В работе М. Р. Когаловского (2000) отмечаются особенности научных коллекций электронных информационных ресурсов. Во-первых, научные коллекции различаются масштабом круга пользователей. В научных электронных библиотеках могут содержаться как коллекции, имеющие национальную и/или ми-

ровую значимость, так и персональные коллекции ученых и исследовательских лабораторий. При этом персональные коллекции со временем могут приобрести более высокий статус. Во-вторых, научные коллекции более динамичны по структуре и составу информационных ресурсов по сравнению, например, с музейными коллекциями. В-третьих, информационные потребности исследователей более разнообразны, чем, например, относительно стабильные информационные потребности управленческих работников. Поэтому в научных ЭБ для доступа к коллекциям должны предусматриваться гибкие пользовательские интерфейсы. В-четвертых, достоверность сведений в научных коллекциях определяется достигнутым уровнем знаний в данной области науки. Также отмечается присущая научным коллекциям неполнота и/или нечеткость представленных ими сведений. При этом научные коллекции могут содержать не только сведения о реальных процессах и явлениях, но и данные компьютерного моделирования с моделями исследуемых систем. Коллекции в различных областях науки имеют свои специфические особенности. Так, например, в географии, экологии и региональной экономике значительную часть информационных ресурсов составляют картографические сведения, для работы с которыми используются ГИС-технологии. Во всех областях научных исследований используются коллекции библиографических данных, а также коллекции полнотекстовых научных публикаций.

В настоящее время у нас в стране и за рубежом накоплен значительный опыт в создании электронных библиотек. Создан портал «Российские электронные библиотеки» – <http://www.elbib.ru>, следует отметить ЭБ РФФИ (<http://elibrary.ru>) и СО РАН (<http://www-sbras.nsc.ru/win/elbib>), а также ряд зарубежных систем – The Historic Pittsburgh Digital Library (<http://www.pitt.edu/~edwardg/histpitt/survey.html>), The Alexandria Digital Library (<http://www.alexandria.ucsb.edu/>) и др. Реализация этих и других проектов создания ЭБ показала практическую значимость и эффективность использования таких информационных систем в первую очередь для поддержки человеческой деятельности в различных областях науки, образования и культуры.

Необходимость создания ЕИС КарНЦ РАН

В Российской академии наук разработана Концепция создания Единой информационной системы (ЕИС РАН), текст которой размещен на информационном Web-портале РАН – <http://www.ras.ru/scientificactivity/eis/eisconception.aspx>.

ЕИС РАН определяется как «...объединение в интегрированное информационное пространство распределенных и локальных цифровых (электронных) ресурсов организаций

РАН (информационных (научных и административных), программных, алгоритмических) и комплекса программно-технических средств, обеспечивающего использование этих ресурсов и полнофункциональное управление ими». При этом отмечается специфическая особенность РАН как объекта информатизации, заключающаяся в том, что значительную часть общекорпоративного информационного потока составляет научная информация, которая является разнородной (тексты, электронные карты, математические формулы, аудио- и видеообъекты и т. п.) и, как правило, слабоструктурированной.

Разработчики программы придают важное значение вопросам интеграции и управления информационными ресурсами. При этом ставятся цели обеспечения возможности доступа всех субъектов РАН к уже имеющимся цифровым научным информационным ресурсам, а также возможности эффективного управления РАН как научной корпорацией. Задача управления ресурсами возлагается на информационно-управляющее ядро системы, представляющее собой совокупность инженерно-технических решений и организационно-технических структур и мероприятий, поддерживающих управление цифровыми ресурсами ЕИС РАН.

В концепции дан анализ текущего состояния информационного пространства РАН. В частности, отмечается, что организации РАН обладают значительными и постоянно возрастающими информационными ресурсами, которые слабо представлены для доступа по телекоммуникационным каналам. Отсутствует централизованная специализированная система поиска научной информации и доступа к ней, а также средства интеграции информационных ресурсов по областям знаний.

Предпосылки информатизации РАН связаны с необходимостью в упорядочении и структуризации доступа к результатам научной деятельности РАН, имеющим распределенный характер. При этом утверждается, что Президиум РАН должен стать организующим центром телекоммуникационного доступа к научной информации. Также отмечается необходимость создания информационно-справочного сервиса, средств интеграции информационных ресурсов, обеспечения доступа к электронным библиотекам РАН.

В качестве обстоятельств, препятствующих созданию единого информационного пространства РАН, отмечается отсутствие понимания необходимости развития работ в этом направлении и как следствие – отсутствие программы работ, неравномерность готовности различных организаций РАН к внедрению и использованию такого рода системы и т. п.

Главные цели проекта создания ЕИС РАН заключаются в том, чтобы обеспечить:

- информационную поддержку организационно-управленческой деятельности научных учреждений и Президиума РАН;

- поддержку научной и экспертно-аналитической деятельности РАН;

- информационное обеспечение учета НИР РАН и поддержку решения задач управления научной корпорацией;

- взаимодействие с внешними информационными системами.

Среди актуальных задач выделяют задачи, связанные с разработкой Единой информационной системы Президиума РАН как компонента ЕИС РАН, создание электронных библиотек, систем автоматизации управленческой деятельности (электронного документооборота), организации взаимодействия с российскими региональными и зарубежными информационными системами, а также с информационными системами РФФИ, РГНФ и др.

ЕИС РАН разрабатывается как распределенная система, состоящая из множества узлов, образующих единое интегрированное информационное пространство. Основу системы должны составлять информационные системы Институты, Отделений и Президиума РАН. При этом информационная система Института РАН должна, с одной стороны, стать центром научно-информационного сервиса сотрудников Института, а с другой – обеспечить представление информации о научной деятельности Института для российского и мирового сообщества. Также отмечается важность наличия средств защиты информации и контроля за доступом к данным системы.

В концепции отмечается, что для разработки ЕИС РАН необходимо стандартизировать форматы метаданных и форматы представления данных, форматы электронного обмена данными для различных областей науки, стандарты именования ресурсов и т. п. Предполагается использовать разнообразные технологии при условии наличия: открытого стандарта; документации на данные и интерфейсы; инструментальных средств разработки.

В настоящее время Карельский научный центр РАН располагает значительными техническими и информационными ресурсами и имеет опыт создания и развития телекоммуникационных и информационных систем. В институтах КарНЦ РАН ведутся работы по созданию и развитию научных электронных информационных ресурсов, в том числе с использованием Internet-технологий. С 1997 г. функционирует официальный Web-сайт Центра (<http://www.krc.karelia.ru>), сайты институтов (<http://biology.krc.karelia.ru>, <http://geoserv.krc.karelia.ru>, <http://nwpi.krc.karelia.ru>), а также ряд тематических предметно-ориентированных сайтов с интегрированными базами данных, например, в области гуманитарных наук – <http://phonogr.krc.karelia.ru>, система TORIS – <http://toris.krc.karelia.ru>, в биологии – <http://biodiv.krc.karelia.ru> и др. (Сорокин и др., 2000). При этом система TORIS включает как базу данных о топонимах, так и программ-

ный комплекс для анализа данных – DMiner, в основе которого лежит алгоритм поиска значимых множеств – PrefixSpan (Вдовицын и др., 2003; Керт, Вдовицын, 2005). С помощью разработанных программ анализа данных исследователи могут находить новые закономерности и осуществлять проверку своих гипотез и предположений, связанных с решением задач в области топонимики. Эта разработка в перспективе может стать основой для создания «виртуальных» предметных лабораторий, в которых ученые из разных регионов могут объединить свои усилия для решения актуальных научных проблем с использованием возможностей современных Интернет-технологий.

Также следует отметить ряд проектов, связанных с публикацией в Интернет результатов комплексных исследований по наиболее значимым природным объектам Карелии, например, по Ладожскому озеру – <http://lado-ga.krc.karelia.ru>, по национальному парку «Койтайоки – Толвоярви» – <http://suotacis.krc.karelia.ru>, электронный каталог Научной библиотеки КарНЦ РАН (<http://foliant.ru/catalog/knclibr/>); работы по созданию электронной коллекции знаний об Онежском и Ладожском озерах (Институт водных проблем Севера и ИПМИ) (Филатов, Лебедев, 2001); проекты в области создания ГИС (Институт геологии).

С 2002 г. ведется работа по созданию и развитию электронной библиотеки научных информационных ресурсов КарНЦ РАН (ЭБ КарНЦ РАН – <http://dl.krc.karelia.ru>), которая поддержана грантами РФФИ № 02-07-90204 и № 05-07-90077. В рамках выполнения этого проекта разработаны и развиваются программные сервисы, обеспечивающие поддержку процессов публикации, поиска и сопровождения цифровых коллекций научных ресурсов.

Большинство перечисленных проектов выполнялись в течение 1997–2005 гг. при поддержке грантов РФФИ, РГНФ, ФЦП «Интеграция» и др. При разработке информационных систем использовались традиционные Web-технологии и свободные распространяемые СУБД – miniSQL и MySQL с организацией доступа к базам данных по разработанным интерфейсным формам с применением технологии CGI-скриптов.

Проведенная в мае 2005 г. инвентаризация электронных научных информационных ресурсов показала, что состояние научного информационного пространства КарНЦ РАН во многом совпадает с оценкой состояния научного информационного пространства РАН. Во-первых, институты Центра обладают значительными и постоянно возрастающими научными информационными ресурсами, которые, как правило, слабо систематизированы и разрознены. Во-вторых, научная информация слабо представлена для доступа по телекоммуникационным каналам. В-третьих, отсутствует централи-

зованная специализированная система поиска научной информации и доступа к ней. В-четвертых, практически отсутствуют для широкого доступа электронные версии изданий научных трудов ученых Центра, хотя авторы в основном предоставляют в издательство электронные варианты публикаций. В-пятых, научная информация не имеет стандартизованного электронного представления.

Таким образом, необходимость создания Единой информационной системы в Карельском научном центре РАН (ЕИС КарНЦ РАН) также во многом перекликается с основными предпосылками информатизации РАН. Во-первых, целесообразно упорядочить и стандартизовать доступ к результатам научной деятельности КарНЦ РАН. Во-вторых, необходимо расширить и реализовать на качественно новом уровне ряд систем автоматизации административной, хозяйственной и научной деятельности Центра, а также организовать электронный обмен научной, хозяйственно-административной и нормативно-правовой информацией между организациями РАН. В-третьих, целесообразно организовать информационно-справочный сервис, осуществляющий накопление и предоставление оперативной научной и хозяйственно-административной информации. В-четвертых, необходимо обеспечить создание коллекции электронных версий научных публикаций сотрудников Центра и доступ к ней по сети Internet. В-пятых, необходимо обеспечить создание электронных библиотек научных информационных ресурсов и доступ к ним по сети Internet, а также доступ к основным мировым электронным базам информационных ресурсов.

Вопросы разработки ЕИС КарНЦ РАН

ОСНОВНЫЕ ТРЕБОВАНИЯ И ПЕРВООЧЕРЕДНЫЕ ЗАДАЧИ

ЕИС КарНЦ РАН должна иметь распределенную архитектуру – состоять из множества серверов, рабочих станций, каналов связи и необходимого телекоммуникационного оборудования, образующих единую интегрированную информационно-телекоммуникационную среду Центра. Основу системы должны составить информационные системы институтов и Президиума КарНЦ РАН.

Разработка и реализация ЕИС КарНЦ РАН должна проводиться с учетом выполнения следующих основных требований:

– с учетом иерархической структуры РАН. Основным учреждением, проводящим научные исследования, является институт, а основными управляющими структурами (в нашем случае) – Президиум КарНЦ РАН, отделения и Президиум РАН. Поэтому главные информационные потоки должны быть направлены как по цепочке «институт – отделение», так и по цепочке «институт – Президиум КарНЦ РАН – Президиум РАН»;

– с учетом тесной интеграции системы с ЕИС РАН. При этом ЕИС КарНЦ РАН должна разрабатываться и развиваться как самостоятельная информационная система (в соответствии со своими целями и потребностями), а с другой стороны, ее разработка и развитие должны осуществляться в тесном взаимодействии с создаваемой ЕИС РАН;

– система должна разрабатываться на основе согласованных стандартов на интерфейсы и метаданные, иметь открытую архитектуру и быть масштабируемой;

– система должна разрабатываться и развиваться поэтапно с учетом появления новых задач и изменения требований к существующим задачам;

– в системе должна быть предусмотрена функция обеспечения охраны авторских прав ученых Центра на научные открытия и изобретения;

– система должна иметь эффективный механизм разграничения прав доступа пользователей к своим ресурсам и обладать надежной защитой информационных ресурсов от несанкционированного доступа. При этом одна часть информации должна быть представлена в открытом, а другая – в строго контролируемом режиме (перечень открытой и информации ограниченного доступа устанавливается на основании распоряжений Президиума РАН, отделений, руководства Центра и институтов).

Первоочередными задачами по созданию ЕИС КарНЦ РАН на ближайшую перспективу, на наш взгляд, являются следующие:

- Развитие и поддержка в актуальном состоянии официального Web-сайта КарНЦ РАН (превращение его в научный портал) и сайтов институтов Центра. Стандартизация их структуры и информационного содержания с учетом фокусировки на информационные потребности целевой аудитории (ученые, преподаватели, инвесторы и т. п.). При этом следует стремиться к тому, чтобы официальный сайт и сайты институтов Центра в большей степени удовлетворяли требованиям современных НИС.

- Создание и развитие тематических (проблемно-ориентированных) Web-сайтов с интегрированными базами данных и знаний, алгоритмическими и вычислительными ресурсами и электронными публикациями по научной тематике, исследования по которой проводятся специалистами Центра. В перспективе при создании таких систем необходимо использовать достижения GRID-технологий.

- Создание инфраструктуры электронной библиотеки научных информационных ресурсов КарНЦ РАН, предназначенной для поддержки процессов публикации, поиска, сохранения и распространения научной информации по сети Интернет.

- Создание и развитие электронных (цифровых) коллекций документов по различным областям знаний, исследования по которым про-

водятся в Центре, и их интеграция в электронной библиотеке научных информационных ресурсов КарНЦ РАН.

- Создание или освоение приобретенных программных комплексов информационной поддержки организационно-управленческой деятельности руководства институтов и Президиума КарНЦ РАН.

- Обеспечение защиты электронных информационных ресурсов Центра от несанкционированного доступа.

- Обеспечение интеграции ЕИС КарНЦ РАН с ЕИС РАН, а также взаимодействие с российскими и зарубежными информационными системами – РФФИ, РГНФ, ГСНТИ, Росинформресурс, российских вузов и др.

В настоящее время в рамках проводимых в Центре работ по созданию элементов ЕИС КарНЦ РАН реализуется ряд проектов, связанных с разработкой информационных систем для поддержки научных исследований и прикладных разработок. В качестве примера можно привести создание информационной системы по топонимии Европейского Севера России (Керт, Вдовицын, 2005), а также электронной библиотеки научных информационных ресурсов КарНЦ РАН.

ЭЛЕКТРОННАЯ БИБЛИОТЕКА КарНЦ РАН КАК ОСНОВА ДЛЯ ИНТЕГРАЦИИ ЭЛЕКТРОННЫХ НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ ЦЕНТРА

Создание электронной библиотеки научных электронных (цифровых) информационных ресурсов необходимо для поддержки процессов проведения фундаментальных исследований и прикладных разработок. Систематизация и структуризация данных и знаний и их представление в виде электронных коллекций документов, построенных по единым согласованным стандартам и технологиям, создают основу для интеграции научных электронных информационных ресурсов Центра, позволяя большому числу российских и зарубежных специалистов получить через Интернет доступ к уникальной научной информации, а также будут способствовать проведению комплексных исследований и решению прикладных задач.

Разработанная нами инфраструктура ЭБ научных информационных ресурсов КарНЦ РАН включает следующие основные сервисы (Вдовицын, Сорокин, 2002, 2003; Вдовицын и др., 2005):

- публикация новой коллекции документов в ЭБ и ее сопровождение (пополнение документами, корректировка и удаление документов);

- поиск информации по запросам пользователя;

- управление доступом к сервисам ЭБ и защита информационных ресурсов.

Структура документов каждой коллекции разрабатывается на основе представленных

специалистами-предметниками паспортов описаний изучаемых объектов и реализуется с помощью языка XML (<http://www.w3c.org/XML/overview.html>). Наиболее привлекательной особенностью XML (по сравнению с HTML) является, в частности, возможность отделения данных от их визуального представления и использование тегов разметки для определения логической структуры текста. Для описания общих свойств каждой коллекции применяется широко распространенный стандарт Дублинского ядра (Dublin Core, DC, <http://purl.org/dc/documents/>). На основе атрибутов стандарта DC создан каталог коллекций, который используется для организации поиска коллекции по запросу пользователя.

Основными функциями программного обеспечения, предназначенного для поддержки процессов публикации и сопровождения документов в ЭБ КарНЦ РАН, являются:

- помощь специалисту-предметнику в заполнении (корректировке) полей документа коллекции в соответствии с разработанной структурой документа – DTD (Document Type Definition) языка XML;
- автоматическое формирование корректного XML-документа коллекции и его сохранение на сервере ЭБ;
- организация на форуме обсуждения экспертами содержания вносимого в коллекцию документа.

Привлечение специалистов-предметников и экспертов к непосредственному участию в процессе создания и сопровождения научной коллекции способствует, на наш взгляд, улучшению качества и достоверности информационного содержания документов и повышает их заинтересованность в конечном результате работы. По существу, каждый документ коллекции можно рассматривать как электронную публикацию, в которой в сжатой форме представлена обобщающая научная информация по результатам многолетних исследований описываемого объекта коллекции.

Процедура поиска информационных ресурсов в ЭБ КарНЦ РАН включает два основных этапа – поиск искомой коллекции среди множества коллекций электронной библиотеки и поиск документов в найденной коллекции по запросу пользователя. Поиск искомой коллекции в ЭБ может быть осуществлен как при помощи рубрикатора ГРНТИ, так и с помощью базы метаданных (каталога), описывающей коллекции на основе атрибутов стандарта DC. Следует отметить, что в систему могут быть интегрированы на уровне своих метаописаний в стандарте DC электронные информационные ресурсы, разработанные по другим информационным технологиям (реляционные базы данных, коллекции электронных карт и т. п.). В этом случае разработанный поисковый сервис ЭБ выступает в качестве «точки входа» в электронное информационное пространство Цент-

ра. Запросы на поиск документов в выбранной коллекции формируются пользователем с помощью специально разработанных интерфейсных форм, специфичных для каждой коллекции.

В настоящее время разработана и развивается первая версия программного обеспечения ЭБ КарНЦ РАН (<http://dl.krc.karelia.ru>). По разработанной технологии созданы и пополняются новыми документами три коллекции научных информационных ресурсов: «Афиллофороидные грибы Карелии» (150 документов), «Виртуальная флора Карелии» (99 документов) и «Млекопитающие Карелии» (31 документ). Основой создания этих электронных коллекций явился большой объем научной информации, полученный сотрудниками институтов биологии и леса Карельского научного центра РАН в результате многолетних исследований.

Развитие ЭБ КарНЦ РАН связано как с созданием новых программных сервисов системы, так и с расширением информационного содержания.

Данный проект поддержан РФФИ (гранты № 02-07-90204, № 05-07-90077).

Заключение

Вопросы формирования и использования электронных научных информационных ресурсов требуют решения целого ряда научно-методических, организационных, финансовых и кадровых вопросов, от которых зависит эффективность создания и функционирования информационной системы. Наиболее значительными из них в нашем случае является разработка и принятие на Президиуме КарНЦ РАН комплексной программы работ по созданию Единой Информационной Системы Центра. При этом разработка и выполнение мероприятий программы должны осуществляться в тесной координации с разработчиками ЕИС РАН.

При разработке и реализации мероприятий программы необходимо учесть проведение исследований с целью выбора методов организации метаданных, интерфейсов, классификаторов и рубрикаторов, перспективных общесистемных информационных технологий и программных средств, предназначенных для построения отдельных информационных систем ЕИС Центра. Также следует разработать технологии администрирования системы, обеспечения информационной безопасности и аутентификации, взаимодействия с управляющим ядром ЕИС РАН. Необходимо решить вопросы защиты авторских и иных прав научных работников на создаваемые электронные информационные ресурсы на основе существующего законодательства РФ.

Реализация мероприятий программы позволит нам достичь следующих основных целей. Во-первых, улучшить доступ научных сотрудников

к необходимым электронным информационным ресурсам. Во-вторых, обеспечить публикацию в Интернет результатов научных исследований. В-третьих, приступить к созданию и применению новых технологий проведения научных исследований в среде Интернет, например, путем создания так называемых «виртуальных» предметных лабораторий с использованием возможностей GRID-технологий. В-четвертых, обеспечить сохранение накопленной научной информации для последующих поколений ученых. В-пятых, обеспечить интеграцию информационной системы Центра в Единое Научное Информационное Пространство Российской академии наук.

Литература

- Армс В., 2001. Электронные библиотеки / Пер. с англ. ПИК ВИНТИ. 274 с.
- Арнаутов С. А., 2001. Роль и место виртуальных цифровых библиотек в Интернете // RCDL'2001: Сб. тр. 3-й Всерос. конф. по Электронным Библиотекам «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Петрозаводск, 11–13 сент. 2001 г.). Петрозаводск: Карельский научный центр РАН. С. 20–25.
- Вдовицын В. Т., Сорокин А. Д., 2002. Вопросы построения электронной библиотеки Карельского научного центра РАН // Тр. 4-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL'2002 (Дубна, 15–17 окт. 2002 г.): В 2 т. Т. 1. Дубна: ОИЯИ. С. 304–308.
- Вдовицын В. Т., Сорокин А. Д., 2003. Технология публикации и сопровождения документов в коллекциях научных информационных ресурсов электронной библиотеки КарНЦ РАН // Тр. 5-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL'2003 (Санкт-Петербург, 29–31 окт. 2003 г.). С. 103–105.
- Вдовицын В. Т., Керт Г. М., Луговая Н. Б., Чуйко Ю. В., 2003. Применение алгоритмов поиска логических зависимостей для решения задач в области топоники // Обозрение прикладной и промышленной математики. Т. 10, вып. 2. С. 387–388.
- Вдовицын В. Т., Сорокин А. Д., Луговая Н. Б., 2005. Развитие программных сервисов и контента ЭБ КарНЦ РАН // Тр. 7-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Ярославль, 4–6 окт. 2005 г.). Ярославль: Ярославский гос. ун-т им. П. Г. Демидова. С. 92–97.
- Жучков А. В., Арнаутов С. А., 2001. Единая среда распределенных ресурсов (GRID) и цифровые библиотеки // RCDL'2001... С. 46–50.
- Ильин В. А., Корежков В. В., 2002. Создание российского сегмента европейской инфраструктуры ЕИ DATAGRID // Тр. 4-й Всерос. науч. конф. «Электронные библиотеки...». Т. 1. С. 239–248.
- Керт Г. М., Вдовицын В. Т., 2005. Информационные технологии в исследовании топонимии // Вопросы языкознания. № 3. С. 102–124.
- Коголовский М. Р., 2000. Систематика коллекций информационных ресурсов в электронных библиотеках // Программирование. № 3. С. 31–52.
- Коголовский М. Р., 2001. Стандарты платформы XML и базы данных // RCDL'2001... С. 108–117.
- Коголовский М. Р., 2002. Энциклопедия технологий баз данных. М.: Финансы и статистика. 800 с.
- Кулагин М. В., Лопатенко А. С., 2001. Научные информационные системы и электронные библиотеки. Потребность в интеграции // RCDL'2001... С. 14–19.
- Лопатенко А. С. Научные информационные системы. Перспективы использования (http://derpi.tuwien.ac.at/~andrei/CRIS_DOC.htm).
- Сорокин А. Д., Вдовицын В. Т., Луговая Н. Б., 2000. Создание и развитие электронных информационных ресурсов в КарНЦ РАН // Сб. докл. 2-й Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Протвино, 26–28 сент. 2000 г.). Протвино. С. 3–5.
- Филатов Н. Н., Лебедев В. А., 2001. Коллекция знаний об Онежском и Ладожском озерах // RCDL'2001... С. 202–208.
- CERIF-2000 Guidelines, Final Report of the CERIF Revision Working Group, Part 1, С. 3 User requirements (<ftp://ftp.cordis.lu/pub/cerif/docs/cerif2000.htm>).