

## ПРОБЛЕМАТИКА ТЕХНОЛОГИЙ ЭЛЕКТРОННЫХ БИБЛИОТЕК<sup>1</sup> НА XI ВСЕРОССИЙСКОЙ НАУЧНОЙ КОНФЕРЕНЦИИ RCDL'2009

*В.Т. Вдовицын, к.ф.-м.н., Карельский научный центр РАН,  
Л.А. Калининченко, д.ф.-м.н., Институт проблем информатики РАН,  
М.Р. Когаловский, к.т.н., Институт проблем рынка РАН,  
С.Д. Кузнецов, д.т.н., Институт системного программирования РАН,  
В.В. Мазалов, д.ф.-м.н., Институт прикладных математических исследований КарНЦ РАН*

С 17 по 21 сентября в Петрозаводске в Карельском научном центре РАН состоялась очередная конференция серии Всероссийских научных конференций «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL'2009. Использование англоязычной аббревиатуры в качестве краткого названия конференции связано с ее англоязычным названием — Russian Conference on Digital Libraries, используемым для информирования о ней международного научного сообщества, представители которого участвуют в работе RCDL и вносят значительный вклад в ее программу.

Конференции RCDL стали проводиться ежегодно при неизменной поддержке РФФИ, после того как Фонд впервые объявил в 1998 г. о поддержке проектов по тематике электронных библиотек и гранты в рамках конкурса РФФИ 1998 г. были предоставлены 29 проектам. За прошедшие годы РФФИ было поддержано более 700 проектов по этой тематике.

Решение об организации серии ежегодных конференций RCDL было принято на состоявшемся по инициативе РФФИ в 1998 году семинаре-совещании первых его грантодержателей по тематике электронных библиотек с тем же названием. Первая конференция проводилась в 1999 году в Санкт-Петербургском государствен-

ном университете. Последующие конференции RCDL состоялись в Протвино (2000), Петрозаводске (2001), Дубне (2002), Санкт-Петербурге (2003), Пушкино (2004), Ярославле (2005), Суздале (2006), Переславле-Залесском (2007), Дубне (2008), Петрозаводске (2009). Такая широкая география мест проведения конференций RCDL по замыслу ее Руководящего комитета должна способствовать вовлечению в разработки и исследования электронных библиотек и технологий их создания более широкого круга специалистов в различных регионах страны. За прошедшие годы в работе конференции приняла участие большая группа ученых из ведущих российских и зарубежных научных центров Австрии, Германии, Греции, Италии, Новой Зеландии, США, Украины, Швеции и других стран, а также специалисты из крупных отечественных компаний, специализирующихся в области информационных технологий.

Традиционно совместно с RCDL проводятся Всероссийские научные семинары по оценке методов текстового поиска — Российский семинар по оценке методов информационного поиска (РОМИП). В 2009 году конференции RCDL'2009 сопутствовало также проведение Третьей Российской летней школы по информационному поиску RuSSIR'2009, во время которой ведущие российские и зарубежные ученые прочитали обзорные лекции по актуальным проблемам развития поиска цифровых данных для решения фундаментальных и прикладных задач. В рамках RCDL'2009 был также впервые организован семинар диссертационных исследований по тематике, связанной с электронными библиотеками.

Труды всех состоявшихся конференций RCDL предоставлены для открытого доступа на сайте конференции [www.rcdl.ru](http://www.rcdl.ru).

Электронные библиотеки — область исследований и разработок, направленных на развитие теории и практики обработки, распространения, хранения, поиска и анализа цифровых данных различной природы. Электронные библиотеки, являющиеся хранилищами знаний, можно рассматривать как сложные информационные системы,



Главное здание Карельского научного центра РАН

<sup>1</sup> Конференция проведена при поддержке РФФИ по гранту № 09-07-06041-г.

при создании и использовании которых требуется решение многих научных, технологических, методологических, экономических, правовых и других вопросов. Развитие технологий электронных библиотек становится всё более существенным для совершенствования стандартов здравоохранения, образования, науки, экономики, равно как и качества жизни вообще. Проекты формирования в цифровой форме информации, накопленной человечеством о Земле, Вселенной, науке, литературе, искусстве, окружающей среде, человеке, являются примерами областей интенсивного развития глобальных репозиториях представления знаний.

Основная цель серии конференций RCDL заключается в том, чтобы способствовать формированию в России сообщества специалистов, ведущих исследования и разработки в области электронных библиотек. Конференция также способствует изучению зарубежного опыта, развитию международного сотрудничества в области электронных библиотек.

Значительное внимание в тематике RCDL уделяется практическим проектам, в которых решаются сложные задачи. При формировании программ RCDL придается большое значение обсуждению исследований в области создания электронных библиотек для поддержки фундаментальных научных исследований, цифровому сохранению научного и культурного наследия, формированию электронных коллекций научных информационных ресурсов в различных областях знаний. Доклады грантодержателей РФФИ по результатам выполняемых ими проектов традиционно составляют значительную часть программы конференций серии. Обсуждение этих работ на конференции существенным образом способствует повышению качества получаемых результатов, ознакомлению научной общественности с актуальными направлениями научных исследований, достигнутыми результатами.

В программу каждой из конференций RCDL традиционно входит проведение тьюториалов, знакомящих слушателей с состоянием исследований в одной из связанных с электронными библиотеками областей информационных технологий. Важной целью RCDL является привлечение молодых ученых и студентов к работе над современными проблемами в этой области; создание площадки для неформального общения представителей науки, высшей школы и индустрии.

Проблематика электронных библиотек довольно обширна. В агрегированном виде она состоит из четырех составляющих: методология разработки электронных библиотек, цифровое сохранение информационных ресурсов и создание контента, сервисы электронных библиотек и их технологическая инфраструктура. В документах конференций RCDL эта проблематика представ-

лена более детально в виде следующего перечня направлений исследований и разработок:

- Методы и модели представления информации.
- Метаданные, онтологии и средства представления знаний.
- Перспективные инфраструктуры электронных библиотек, электронные библиотеки в средах грид и облачного компьютеринга.
- Виртуальные среды для научных исследований.
- Социальные сети и электронные библиотеки.
- Инструментальные средства создания электронных библиотек и цифровых коллекций.
- Методы и технологии работы с текстовой, мультимедийной и слабоструктурированной информацией.
- Методы сохранения информации и цифровые архивы.
- Интероперабельность и интеграция информации.
- Наукометрические измерения в электронных библиотеках.
- Информационная безопасность.
- Пользовательские интерфейсы, человеческий фактор.
- Электронные библиотеки в науке, образовании, культуре, медицине, экономике.
- Стандарты технологий электронных библиотек.

В результате рецензирования поступивших 68 расширенных тезисов Программный комитет конференции RCDL'2009 отобрал для включения в программу конференции 32 полных, 19 кратких и 3 стендовых доклада. Около 40% из них было представлено грантодержателями РФФИ. Кроме того, в программу конференции было включено два приглашенных доклада и тьюториал. На этой основе была сформирована научная программа конференции.

Конференции цикла «Электронные библиотеки: перспективные методы и технологии, электронные коллекции — RCDL» являются в нашей стране единственным ежегодным научным форумом национального масштаба, специально посвященным проблематике информационных технологий, используемых в электронных библиотеках. Поэтому включенные в программы этих конференций доклады отечественных авторов позволяют получить достаточно полное представление о направлениях исследований в области электронных библиотек, проводимых отечественными учеными, и о научном уровне достигнутых результатов.

Нужно отметить, что многие пласты информационных технологий, которые применяются в электронных библиотеках, составляющих осо-

бый класс информационных систем (ИС), имеют значительно более широкую сферу применения в разработках ИС различного назначения и не предназначены специально для создания электронных библиотек. Таким образом, новые достижения в таких информационных технологиях активно используются в электронных библиотеках. С другой стороны, развитие исследований и разработок в области электронных библиотек существенным образом стимулирует прогресс в сфере таких общих технологий.

Исследования и разработки, представленные на конференции RCDL'2009, проводятся отечественными исследователями в целом ряде актуальных направлений. Главные из них связаны с развитием эффективных методов разработки методов и технологий для обеспечения новых функциональных возможностей систем электронных библиотек, создания комфортных для пользователей средств доступа к коллекциям информационных ресурсов, надежного сохранения научного, образовательного и культурного наследия для ныне живущих и будущих поколений.

К числу указанных направлений относится, прежде всего, обсуждаемая в ряде представленных на конференции докладов проблема цифрового сохранения. Фундаментальным контекстом для обсуждения этой проблематики послужил приглашенный доклад известного австрийского специалиста в данной области А. Раубера. Отечественные работы этого направления посвящены созданию конкретных научно-образовательных электронных библиотек, а также разработкам электронных коллекций цифровых информационных ресурсов в различных областях науки и в области образования. В докладах на конференции, в частности, были представлены разработки электронных библиотек для научно-исследовательских учреждений и вузов, библиотек для отдельных научных направлений, инструментальных средств для создания цифровых архивов и крупных систем управления контентом, научных цифровых коллекций информационных ресурсов в области астрономии, ядерной физики, физики атмосферы, биологии и в других областях знаний.

Многолетнюю историю имеет другое направление исследований, которое было представлено на конференции и посвящено решению сложной проблемы семантической интеграции неоднородных информационных ресурсов из множества источников. Разработанный отечественными учеными подход и реализующий его инструментарий позволяют создавать конкретные системы интеграции данных. На предыдущих конференциях RCDL докладывались результаты разработки прототипов систем интеграции в области молекулярной биологии и астрономии, основанное на предложенном подходе. На конференции



Заседание секции конференции

RCDL'2009 были представлены новые результаты в этой области, а также обзорный доклад по проблемам семантической интеграции данных.

Обеспечение поддержки в электронных библиотеках семантики информационных ресурсов и создание на этой основе пользовательских интерфейсов семантического уровня — весьма актуальная проблема в области электронных библиотек. В этом направлении важное место занимает онтологическое моделирование предметной области. Хотя в настоящее время уже появились коммерческие и свободно распространяемые инструментальные средства инженерии онтологий (языки описания онтологий и программные средства), в этой области все еще остаются нерешенные проблемы, возникающие, в частности, при решении задачи интеграции информационных ресурсов. Кроме того, в международном научном сообществе, к сожалению, пока еще не сложилось ясного понимания соотношения между характером онтологического и концептуального моделирования, в связи с чем часто возникают коллизии. Методологическим проблемам онтологического моделирования как общего характера, так и в приложении к конкретным областям научных исследований на конференции был посвящен ряд докладов. Здесь были представлены некоторые новые результаты.

В программе конференции нашли также отражение активно проводимые в нашей стране исследования в области обработки и поиска текстовых информационных ресурсов. Коллекции текстов на естественных языках составляют весьма значительную долю информационных ресурсов электронных библиотек. Поэтому данное направление исследований имеет особую значимость в разработках электронных библиотек. Исследования в этой области проводятся не только академическим и вузовским сообществом. Весомый вклад в развитие этого направления принадлежит отечественным коммерческим

компаниям, поставляющим лингвистические услуги и инструментарий пользователям. В данном направлении на конференции был представлен ряд докладов, посвященных методам извлечения информации из текстов, выявлению неявных связей между упоминаемыми в тексте объектами, устранению лексической многозначности терминов, выявлению дубликатов библиографических источников, тематическому упорядочению текстов при формировании новостных лент, опознаванию машинно-сгенерированных текстов, созданию цифрового корпуса текстов с синтаксической разметкой, использованию технологий электронных библиотек в системах машинного перевода.

Электронные библиотеки представляют собой класс информационных систем, функционирующих и доступных в среде Всемирной паутины. Не случайно поэтому, что веб-технологии находят широкое применение в разработках электронных библиотек. Представленные на конференции новые результаты в этой области связаны с созданием новых методов обнаружения поискового спама, исследованием роли веб-сайтов как коммуникаторов, с поддержкой социальных сетей в среде научного сообщества. Обеспечиваемые новыми веб-технологиями возможности поддержки в среде Веб социальных сетей позволяют использовать новую технологию электронных научных публикаций, основанную на концепции «живого» документа. Зарубежные ученые только начинают заниматься этой проблемой. Результаты ее отечественных исследований также обсуждались на конференции. В ряде представленных на конференции разработок используются стандарты семантического Веб.

Среди методологических проблем создания электронных библиотек на конференции RCDL'2009 впервые прозвучала тема концептуального моделирования систем электронных библиотек. Это сравнительно новое направление исследований в данной области. На конференции были представлены в основном обзор известных подходов к решению этой проблемы, обсуждались разрабатываемые для указанной цели стандарты концептуального моделирования.

Как и в любой информационной системе, в электронных библиотеках весьма злободневна проблема информационной безопасности. Она не была обойдена на конференции. В частности, обсуждалась проблема защиты информационных ресурсов электронных библиотек от несанкционированного копирования.

Важная тема прозвучала в приглашенном докладе С.И. Паринова. Автор аргументированно показал, что активное развитие электронных библиотек и поддержка международных инициатив открытого доступа к результатам научных исследований в научном сообществе создают

условия для новой более эффективной формы научно-исследовательской деятельности, которая была названа им открытой наукой. Концепция открытой науки была основательно проанализирована в докладе.

Ряд докладов в программе конференции был посвящен изучению человеческого фактора в системах электронных библиотек. К их числу относятся работы, посвященные изучению пользовательских предпочтений в среде конкретных электронных библиотек, поиску информации в электронных библиотеках с учетом определенного пользователем контекста, анализу возможностей обеспечения предоставления информационных ресурсов по требованию пользователя. Последняя проблема более основательно анализировалась в тьюториале шведского гостя конференции К. Сендкюля.

Отечественные исследования в области электронных библиотек относятся в значительной их части к перспективным направлениям научных исследований связанной с ними проблематики, проводимых международным научным сообществом и охватывающих проблемы как методологии создания электронных библиотек и цифрового сохранения, так и создания инструментария для разработки, сервисов и контента таких систем. По нашему мнению, большинство выполняемых в рассмотренных областях исследований отечественных ученых соответствует мировому уровню научных достижений в рассматриваемой области.

В значительной степени высокий уровень этой сферы отечественных научных исследований связан с влиянием конференций RCDL, которые проводятся в координации с ежегодной научной Европейской конференцией по электронным библиотекам ECDL (European Conference on Digital Libraries). Это способствует возможности приглашения ведущих зарубежных специалистов в области электронных библиотек для выступления с приглашенными докладами и проведения тьюториалов на RCDL по актуальным проблемам исследований и разработок технологий электронных библиотек. Контакты с организаторами ECDL развиваются. Есть надежда, что в ближайшие годы конференция цикла ECDL может состояться в России. Такая возможность в настоящее время изучается. Проведение столь авторитетной и масштабной международной конференции в нашей стране могло бы послужить серьезным стимулирующим фактором для развития исследований в этой области в России.

Вместе с тем имеющийся в настоящее время отечественный научный кадровый потенциал явно не соответствует масштабам этой актуальной проблематики исследований, в значительной степени определяющих степень развития российских элек-

тронных библиотек как важнейшей составляющей инфраструктуры информационного общества в нашей стране и существенного средства поддержки фундаментальных научных исследований. Весьма значительную роль в развитии этой ветви информатики и ее приложений для поддержки фундаментальных исследований в различных областях знаний сыграла объявленная РФФИ в 1998 г. программа «Электронные библиотеки». Наиболее значимые научные результаты по технологиям и методологии разработки электронных библиотек получены именно в рамках проектов, поддерживаемых РФФИ. Об этом говорит тот факт, что большая доля докладов в программе RCDL'2009, как и в программах предыдущих конференций цикла, посвящена обсуждению результатов, полученных в рамках проектов РФФИ. Кроме того, организация и проведение конференций RCDL стали возможны только благодаря поддержке фонда.

В рамках конференции RCDL'2009 впервые в цикле был организован семинар молодых ученых «Диссертационные исследования по тематике информационных технологий, связанных с электронными библиотеками». Цель семинара – предоставить диссертантам возможность обсудить с более опытными коллегами текущие результаты и идеи их исследований; их слабые и сильные стороны, а также направления дальнейшего развития проводимых исследований, установить контакты для научного сотрудничества, развить навыки выступлений на научных конференциях. Руководящий комитет конференций RCDL считает, что такой семинар может служить важным средством активизации исследований молодых ученых в области технологий электронных библиотек и повышению их качества.

Доклады, представленные на семинаре, были тщательно отрецензированы членами специально созданного программного комитета, и в окончательных вариантах докладов, опубликованных в трудах конференции, были учтены замечания

и предложения рецензентов. Тематика докладов вполне соответствовала общей направленности конференции: методы поиска текстовых документов, в том числе основанные на учете семантики текстов; технологии и модели построения тематически ориентированных электронных библиотек и текстовых коллекций; современные средства публикации документов в веб-среде. В целом докладчикам удалось подготовить зрелые и интересные сообщения и вполне удачно представить их участникам семинара.

Хотя программа первого семинара включала всего семь докладов, важно было создать прецедент и отработать технологию проведения такого мероприятия. В семинаре помимо докладчиков приняли участие многие участники конференции, что позволило серьезно обсудить представленные работы. Эффект семинара оценивается участниками как несомненно положительный. Планируется проводить такие семинары на последующих конференциях RCDL.

#### **Решение конференции**

Признано целесообразным найти возможности для публикации лучших докладов будущих конференций RCDL в журналах из списка ВАК, продолжить на следующих конференциях практику организации диссертационного семинара. С целью расширения представительности программного комитета и повышения качества рецензирования заявляемых докладов рекомендовано расширить состав Программного комитета следующей конференции — RCDL'2010, были рекомендованы конкретные кандидатуры. Решено также принять предложение представителей Казанского научного центра РАН, Казанского государственного университета и Научно-исследовательского института математики и механики имени Н.Г. Чеботарева и провести двенадцатую конференцию RCDL'2010 в октябре 2010 г. в Казани на базе Казанского государственного университета. Постоянный сайт конференции находится по адресу <http://rcdl.ru>.

*Все фотографии, использованные в статье, предоставлены авторами.*