

Информационно-аналитическая система поддержки и сопровождения научных исследований природных ресурсов региона*

© А.Ф. Титов, В.Т. Вдовицын, В.А. Лебедев, А.К. Полин

Карельский научный центр РАН, г. Петрозаводск

krcras@krc.karelia.ru

Аннотация

Представлены основные подходы к созданию информационно-аналитической системы поддержки и сопровождения научных исследований природных ресурсов региона. Основное внимание уделено вопросам разработки комплексной онтологии по естественным наукам, охватывающим предметные области исследований участников проекта, которая положена нами в основу технологий систематизации и поиска научной информации. Также представлено описание картографического сервиса, разработанного на основе пакета MapServer v.4.10.3 и стандартов Open Geospatial Consortium с использованием программной оболочки P.MAPPER v.3.2. и технологии AJAX.

1 Введение

Исследования и разработки в области создания информационных систем для поддержки проведения мульти- и междисциплинарных научных исследований ведутся в последнее время как у нас в стране, так и за рубежом. В первую очередь здесь необходимо отметить проект «Электронная Земля: научные информационные ресурсы и информационно-коммуникационные технологии» [14], а также соответствующие проекты по Программе фундаментальных исследований СО РАН [12]. В качестве конкретного примера подобного рода систем можно назвать распределенную информационно-аналитическую систему с единой точкой доступа к геоданным и с инструментарием для их обработки и визуализации [8], а также работы группы специалистов из Томска [13]. Следует подчеркнуть, что во всех этих информационных системах активно используются, прежде всего, современные ГИС и веб-технологии.

Труды 12^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

2 Цель создания информационно-аналитической системы

Научные учреждения (институты) Карельского научного центра РАН и их подразделения на протяжении многих лет проводят всесторонние фундаментальные исследования природных комплексов Восточной Фенноскандии. При этом используются методы математического моделирования, информационно-телекоммуникационные и ГИС-технологии.

В результате этих исследований накоплена очень большая и уникальная научная информация, значительная часть которой может быть представлена в виде электронных коллекций в интернете. В качестве отдельных примеров такого рода локальных баз данных (БД) можно привести: БД «Местообитания Восточной Фенноскандии» (кол-во описаний – 1100, объем – 195 Мбайт) – Институт леса и Институт биологии КарНЦ РАН); БД «Паразиты животных и растений Карелии» (представлены списки более 800 видов паразитов, объем ресурса – 5,5 Мбайт, Институт биологии КарНЦ РАН); БД «Список видов животных Карелии» (представлено 510 таксонов для позвоночных и 1800 таксонов для беспозвоночных животных, объем – 0,3 Мбайт), БД «Перепончатокрылые насекомые Карелии» (в списке – 2387 видов, объем – 7,5 Мбайт) – Институт леса КарНЦ РАН; БД «Зоопланктон озер Карелии» (объем – 15 Мбайт), БД «Гидрофизические характеристики Онежского озера» (объем – 85 Мбайт), «Белое море и водосбор» (объем – 5 Гбайт), «Атлас Онежское озеро» (объем – 4 Гбайт) – Институт водных проблем Севера КарНЦ РАН и др.

Частично эта уникальная научная информация уже представлена в виде цифровых коллекций (например, на портале КарНЦ РАН – <http://www.krc.karelia.ru>, в электронной библиотеке – <http://dl.krc.karelia.ru>), а также локальных БД и отдельных интернет- и ГИС-проектов (например, <http://lakemodel.net>, «ГИС – водные объекты Республики Карелия») [2, 6, 7, 11]. Однако большая ее часть (включая значительное количество картографического материала) носит разрозненный характер, требует систематизации, структурирования и перевода в цифровой формат.

Создание информационно-аналитической системы (ИАС) поддержки и сопровождения научной, аналитической и управленческой деятельности необходимо и важно в первую очередь для эффективной координации и осуществления комплексных (мульти- и междисциплинарных) научных исследований, выполняемых институтами КарНЦ РАН в рамках задач инвентаризации природных ресурсов, при оценке состояния окружающей среды и экологических последствий планируемых и проводимых на территории Карелии и сопредельных регионов мероприятий в сфере промышленности, лесного, сельского и рыбного хозяйства, а также для более успешной интеграции российской науки в мировое научное пространство. Создание такой системы позволит представить в сети (интранет/интернет) актуализированную электронную научную информацию: по геологии, полезным ископаемым; гидрографии, гидрохимии и гидробиологии озер, рек и Белого моря; биологическим ресурсам, о влиянии климатических и антропогенных факторов на биоразнообразие и др. Представление этих научных материалов в виде единой (комплексной) информационной системы будет способствовать (в т. ч. на основе картографического моделирования) научно обоснованному использованию лесов, земель, водоемов, болот, месторождений полезных ископаемых, целенаправленному и планомерному развитию населенных пунктов и транспортных сетей, учету особенностей ландшафтной структуры растительности, мониторингу изменения растительного покрова по данным дистанционного зондирования и т. п.

В конечном итоге на основе многоуровневого мониторинга и предсказательного моделирования с учетом пространственного распределения исследуемых объектов будет создана научная основа для рационального использования природных ресурсов и развития региона.

3 Основные подходы к созданию информационно-аналитической системы

Систематизация и структуризация электронной научной информации, полученной сотрудниками институтов КарНЦ РАН, а также разработка системы информационного поиска осуществляются с применением методов онтологического моделирования. Различные аспекты применения методов онтологического моделирования в информационных системах (в т. ч. для организации информационного поиска) рассматривалось в докладах Симпозиума «Онтологическое моделирование» [10]. Среди известных нам русскоязычных онтологий по естественным наукам, применяемых для индексации научных публикаций, необходимо отметить разрабатываемый в НИВЦ МГУ под руководством Б.В. Доброва проект лингвистической онтологии [5].

В рамках данного проекта для построения и развития комплексной предметной онтологии привлечены ведущие специалисты КарНЦ РАН в области

биологии, сельского и лесного хозяйства, геологии, гидрологии и др. Разрабатываемая онтология положена нами в основу систематизации и интеграции электронных научных информационных ресурсов в единую комплексную информационно-аналитическую систему поддержки и сопровождения научных исследований КарНЦ РАН, а также используется для разработки системы информационного поиска. На базе этой онтологии в настоящее время разработаны и реализованы технологии: предметизации публикаций по отношению к рубрикам Государственного рубрикатора научно-технической информации (ГРНТИ) с формированием соответствующего электронного каталога (систематизация по ГРНТИ); индексации публикаций с формированием базы данных индексов; формирования тематического запроса на поиск документов с использованием онтологии; поиска документов как по рубрикам ГРНТИ, так и по сформированному логическому условию отбора, а также технологии загрузки и редактирования онтологии [3, 4].

Основные номенклатуры терминов и названий объектов, отображенных на карте и описанных в соответствующих атрибутивных базах данных, представлены ниже:

- коренная геология, типы геологических тел, горные породы с разновидностями, геохронологическая шкала;
- тектоника, тектонические циклы, пликативные структуры, тип складки, вид складки, форма складки, дизъюнктивные структуры, вид разрыва, разлома, тип надвига, сброса;
- четвертичная геология, типы и формы рельефа, четвертичные отложения;
- минерально-сырьевые ресурсы, металлические полезные ископаемые, неметаллические полезные ископаемые;
- гидрография, названия рек, озер, водохранилищ, каналов, списки характеристик;
- почвы, типы и виды почв, материнские породы, биопродуктивность;
- ландшафты, типы ландшафтов;
- болота, названия болот и систем, типы болотных участков, массивов и болотных систем;
- леса, типы лесов, типы лесных ландшафтов, лесные биогеоценозы;
- биологические ресурсы, виды ресурсов: ресурсы растительного мира, ресурсы животного мира, характеристики;
- особо охраняемые природные территории (ООПТ), названия ООПТ, категории ООПТ;
- экологический и научный туризм, виды туризма, список маршрутов, список турбаз;
- промышленность и экономическое районирование, экономические районы, отрасли промышленности, крупнейшие предприятия;
- населенные пункты;
- дороги.

Онтология состоит из двух частей: таксономии терминов и совокупности логических условий предметизации электронных публикаций.

Для осуществления эффективного тематического поиска релевантных записей в атрибутивных базах данных и коллекциях электронных научных публикаций с использованием онтологии выполняется их систематизация по отношению к рубрике ГРНТИ, которая осуществляется в два этапа. На первом этапе выполняется анализ метаданных публикации: заголовка и, возможно, списка ключевых слов (если он имеется). При этом выявленные термины сопоставляются с заданным условием и в положительном случае определяется рубрика ГРНТИ, соответствующая публикации (при этом публикация может соответствовать более, чем одной рубрике). Рубрика определяет фрагмент таксономии терминов, по которому будет выполнена индексация данной публикации. Алгоритм индексации осуществляет сканирование текста статьи или описания записи базы данных, выявляет термины и названия, имеющиеся в онтологии, проверяет их соответствие определенному фрагменту онтологии и формирует индекс (все термины и названия на русском языке). Индексы формируются на каждую запись базы данных, содержащую название объекта, которое соответствует названию в онтологии.

Таким образом, в результате индексации электронных научных материалов в ИАС образуется база индексов, в которой впоследствии осуществляется поиск. При этом база индексов упорядочивается по рубрикам ГРНТИ, и тем самым осуществляется систематизация электронных научных материалов.

Поиск необходимой пользователю информации в ИАС может быть осуществлен (помимо обращения к цифровой карте) по некоторому условию, содержащему тематические термины в их логической связи. Построить такое условие не просто. Поэтому предусмотрен сервис, помогающий пользователю формировать соответствующие тематические запросы с использованием онтологии. При этом пользователь сначала выбирает требуемую рубрику рубрикатора ГРНТИ, после чего ему визуализируется соответствующий раздел таксономии, в котором он должен отметить интересующие его термины. Этот список терминов на следующем этапе используется для построения запроса.

В настоящее время ГИС- и интернет-технологии находят широкое применение в различных странах. При этом используется как сравнительно дорогое проприетарное программное обеспечение (ПО), так и свободно-распространяемое. Как правило, страны, использующие проприетарное ПО, имеют и весьма отлаженную инфраструктуру пространственных данных (ИПД) [13], позволяющую в режиме online получать недостающую и актуализированную пространственную информацию (например, картографический сервер Геологической службы Финляндии – <http://geomaps2.gtk.fi/activemap/>). К сожалению, такой путь в настоящее время для большинства научных учреждений РАН весьма затруднен из-за недостаточного финансирования, неразвитости ИПД, различного рода режимных ограничений и т. д.).

Тем не менее, выбор, разработка и использование экономически обоснованной технологии позволяют уже сейчас ввести в практику мульти- и междисциплинарных научных исследований разработанные геоинформационные ресурсы. Такая технологическая схема уже использовалась в КарНЦ РАН при реализации пилотных проектов «Зеленый пояс Фенноскандии» и «Потенциальные объекты экологического туризма РК» и показала высокую эффективность, особенно в отношении малобюджетных проектов [1].

Для создания картографического сервиса нами используется свободно распространяемое программное обеспечение – Linux, Apache v.2.2.4, MySQL v.5.0.45, PHP v.5.2.3. Картографический сервис ИАС разработан на основе пакета MapServer v.4.10.3 и стандартов Open Geospatial Consortium с использованием на первом этапе программной оболочки P.MAPPER v.3.2. и технологии AJAX. MapServer в настоящее время является наиболее популярным, весьма скоростным и функциональным инструментом для создания картографических веб-сервисов, не уступая аналогичному коммерческому ПО. Это экономически обоснованное технологическое решение позволило отказаться от ранее применяемой нами технологии ASP-кар [9]. Открытость программных кодов позволяет наращивать функциональность MapServer. На основе оболочки P.MAPPER нами были разработаны дополнительные сервисы, позволяющие, в частности, работать с всплывающими окнами, уделяя основное внимание собственно карте.

С помощью разработанных ГИС-сервисов проведена работа по формированию и загрузке в ИАС необходимых тематических слоев и соответствующей атрибутивной информации. К настоящему времени на картографический сервер ИАС загружены следующие тематические слои (масштаб М1:1000000), сформированные специалистами-предметниками как из научных учреждений КарНЦ РАН, так и специалистами сторонних организаций, а именно: коренная геология; тектоника; четвертичная геология; полезные ископаемые; геологические памятники; почвенная карта; болота; водные ресурсы (кадастр рек, водосборы 1 и 2 порядков); биотопы; топография (озера, реки, транспортная сеть, населенные пункты, административные границы).

Таким образом, главной особенностью нашей системы (наряду с применением современных ГИС- и веб-технологий) являются создание и использование онтологии для систематизации и структуризации разнородной научной информации и разработка на ее основе эффективной системы информационного поиска. При этом предполагается, что по запросу пользователя система сможет предоставить ему не только соответствующие тематические слои и атрибутивную информацию из баз данных, но и другую сопутствующую научную информацию (электронные научные публикации по теме запроса, документы из электронных библиотек и т. п.), релевантные его информационным потреб-

ностям. Для этого разработана оригинальная технология индексации электронных научных информационных ресурсов, созданных в институтах КарНЦ РАН и других научных организациях. Такой подход к созданию ИАС даст возможность специалистам оперативно анализировать большие объемы разнородной научной информации. Тем самым мы стремимся добиться повышения уровня информационного обеспечения исследователей как при проведении мульти- и междисциплинарных исследований природных ресурсов и окружающей среды региона, так и при осуществлении различного рода комплексных научных эколого-социально-экономических экспертиз народно-хозяйственных программ и проектов.

4 Заключение

Поставленные в рамках данного проекта задачи и подходы к их решению являются, на наш взгляд, актуальными и представляют как теоретический, так и практический интерес. Во-первых, создана интегрированная база научных данных по природным ресурсам Карелии, включающая картографические материалы, атрибутивные базы данных, электронные научные публикации и коллекции. Во-вторых, в качестве объединяющей структуры применена комплексная онтология по естественным наукам, которая разработана совместно со специалистами-предметниками из научных учреждений КарНЦ РАН и охватывает предметные области исследований участников проекта. Эта онтология является основой разработанных технологий систематизации контента и поиска данных, позволяющих пользователю получать по запросу комплекс взаимосвязанной научной информации, релевантный его информационным потребностям. Такой подход к созданию ИАС дает возможность специалистам оперативно анализировать большие объемы разнородной научной информации. Тем самым мы стремимся добиться повышения уровня информационного обеспечения исследователей как при проведении мульти- и междисциплинарных исследований природных ресурсов и окружающей среды региона, так и при осуществлении комплексных научных эколого-социально-экономических экспертиз народно-хозяйственных программ и проектов. Применение при создании ИАС (наряду с современными ГИС- и веб-технологиями) методов онтологического моделирования должно способствовать, по нашему мнению, интеграции разнородной научной информации в единую информационную систему поддержки и сопровождения мульти- и междисциплинарных исследований природных ресурсов региона не только ученых КарНЦ РАН, но и их российских и зарубежных коллег.

Демонстрационный прототип системы представлен в интернете – <http://ias.krc.karelia.ru>. В полном объеме возможности ИАС будут доступны пользователям корпоративной сети интранет научных учреждений КарНЦ РАН.

Работы по созданию системы поддержаны грантами РФФИ № 09-07-12074 офи_м и № 08-07-00085а. Авторы статьи выражают свою благодарность разработчикам основных программных сервисов ИАС – Луговой Н.Б., Гушкалову М., Старковой В.Г., а также всем сотрудникам научных учреждений КарНЦ РАН, участвующим в формировании комплексной онтологии, соответствующих тематических слоев и атрибутивной информации.

Литература

- [1] Кошкарев А.В. Инфраструктура пространственных данных Финляндии //Пространственные данные. – 2008. – № 1. – <http://www.gisa.ru/44536.html>.
- [2] Антипин В.К., Токарев П.Н. Разработка методики составления электронных картографических баз данных растительных ресурсов болот Карелии // Труды Карельского научного центра РАН. Серия Биогеография. – 2008. – Вып. 12. – Петрозаводск, 2008. – С. 3-8.
- [3] Вдовицын В.Т., Лебедев В.А., Луговая Н.Б., Сорокин А.Д., Старкова В.Г. Разработка и развитие технологии публикации и поиска документов в электронных коллекциях. //Труды Восьмой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Суздаль, Россия 17–19 октября 2006 г. – Ярославль, 2006. – С. 162-167.
- [4] Вдовицын В.Т., Лебедев В.А. Онтологии для тематического поиска данных в коллекциях электронной библиотеки. //Труды Десятой Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Дубна, Россия, 7–11 октября 2008. – С. 63-69.
- [5] Добров Б.В., Лукашевич Н.В. и др. Разработка лингвистической онтологии по естественным наукам для решения задач информационного поиска //Труды 7 Всерос. науч. конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL'2005, Ярославль, Россия, 2005.
- [6] Литинский П.Ю. Трехмерное моделирование структуры и динамики таежных ландшафтов. – Петрозаводск: Карельский научный центр РАН, 2007. – 107 с.
- [7] Меншуткин В.В., Филатов Н.Н., Потахин М.С. Разработка экспертной системы «Озера Карелии». Порядковые и номинальные характеристики озер// Водные ресурсы. – 2009. – Т. 2, № 36. – С. 160-171.
- [8] Молородов Ю.И., Смирнов В.В., Федотов А.М. Сервисы геоинформационной системы сбора, хранения и обработки данных натурных наблюдений // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XI Всерос. науч. конф. RCDL'2009. – Петрозаводск: КарНЦ РАН, 2009. – С. 419-424.

- [9] Полин А.К., Скворцова А.С. Применение ГИС и Интернет-технологий для информационного обеспечения процессов рационального природопользования Республики Карелия //Проблемы рационального использования природного и техногенного сырья Баренцева региона в технологии строительных и технических материалов. Материалы II Межд. науч. конф. – Петрозаводск, 2005. – С. 170-172.
- [10] Труды Симпозиума «Онтологическое моделирование». //Под ред. Л.А. Калиниченко. – М.: ИПИ РАН, 2008. – 303 с.
- [11] Филатов Н.Н., Меншуткин В.В. Разработка геоинформационной и экспертной систем для оценки водных и биологических ресурсов водных объектов Карелии// Информационный бюллетень ГИС-Ассоциации. – 2008. – № 4 (66). – С. 1-18.
- [12] Шокин Ю.И., Добрецов Н.Н., Пестунов И.А., Молородов Ю.И., Смирнов В.В., Синявский Ю.Н. Система сбора, хранения и обработки спутниковых и наземных данных Новосибирского научного центра СО РАН // Вычисл. технол. – 2008. – Т. 13. – Вестн. КазНУ им. Аль-Фараби. Серия: Математика, механика, информатика. – № 4 (59). – Совместный вып. по материалам межд. конф. «Вычислительные и информационные технологии в науке, технике и образовании». – Ч. III. – С. 371-376.
- [13] Titov A., Gordov E., Okladnikov I., Shulgina T. Web-system for processing and visualization of meteorological data for Siberian environment research //Int. J. of Digital Earth. – 2009. – V. 2, Issue S1. – P. 105-119.
- [14] http://www.scgis.ru/geosinet/geosinet_info.html

**Information-analytical system
for supporting and maintenance of
scientific research of natural resources
in the region**

A.F. Titov, V.V. Vdovitsyn, V.A. Lebedev, A.K. Polin

The article presents the main approaches to the creation of information-analytical system of support and maintenance of regional natural resources' scientific research. It focuses on developing the comprehensive ontology of the natural sciences, covering the subject areas of project participants' researches, which was chosen as the basis for the technology of scientific data's systematization and search. The article also contains the description of the map service of information and analysis system, developed on the basis of MapServer v.4.10.3 and Open Geospatial Consortium standards with usage of P. MAPPER v.3.2. program shell and AJAX technology.

*Работа выполнена при финансовой поддержке РФФИ (проекты 09-07-12074 офи_м и 08-07-00085)