

СИСТЕМАТИЗАЦИЯ И ПОИСК ДАННЫХ С ПРИМЕНЕНИЕМ ОНТОЛОГИИ В ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СИСТЕМЕ КАРНЦ РАН

В.Т. Вдовицын, В.А. Лебедев, Н.Б. Луговая, В.Г. Старкова

Научные учреждения (институты) Карельского научного центра РАН на протяжении более 50-лет проводят всесторонние фундаментальные исследования природных комплексов Восточной Фенноскандии. В результате этих исследований накоплена уникальная научная информация, которая частично представлена в виде цифровых коллекций (например, на портале КарНЦ РАН – <http://www.krc.karelia.ru>, в электронной библиотеке научных информационных ресурсов - <http://dl.krc.karelia.ru> и др.). Однако большая ее часть (включая значительное количество картографического материала) носит разрозненный характер, требует систематизации, структуризации и перевода в цифровой формат.

Создание информационно-аналитической системы (ИАС) поддержки и сопровождения научной деятельности по природным ресурсам и окружающей среде Карелии необходимо и важно в первую очередь для координации и проведения 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. Представление в сети (Интранет/Интернет) накопленных научных материалов (по геологии, проявлениям полезных ископаемых, гидрографии и гидрологии рек, озер и Белого моря, влиянии климатических и антропогенных факторов на изменение трофности и загрязнения, биоразнообразия и др.) в виде комплексной информационной системы будет способствовать (в т.ч. и 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 (леса и болота), мониторингу изменения растительности по данным дистанционного зондирования и т.п.

Таким образом, одной из особенностей создаваемой ИАС "Природные ресурсы Карелии" является комплектование ее информационной базы разнородными данными – географическими картами, базами данных, электронными коллекциями и электронными научными публикациями. При этом возникает проблема интеграции и систематизации разнородных электронных информационных ресурсов с целью организации эффективного поиска и формирования по запросу пользователя комплекса взаимосвязанной научной информации для обеспечения конкретных исследований и разработок.

Разработка ИАС "Природные ресурсы Карелии" осуществляется с применением современных Web- и ГИС-технологий, а также методов онтологического моделирования [1,2].

Для построения и развития комплексной предметной онтологии в рамках данного проекта привлечены высококвалифицированные специалисты из научных учреждений КарНЦ РАН в области биологии, сельского и лесного хозяйства, геологии, гидрологии и др. В нашем понимании онтология состоит из двух основных частей: таксономии терминов предметной области и совокупности их функциональных связей, необходимых для решения задач, стоящих перед информационной системой (в нашем случае – повышение эффективности информационного поиска посредством систематизации электронной информации при помощи онтологии). Таксономия в данном случае содержит номенклатуры и связи терминов научных дисциплин, а именно, терминов описания их предметов изучения: строения (морфология и анатомия), функционирования (физиология, динамика), взаимодействие с окружающей средой (экология), происхождение, возникновение (генезис). Научные дисциплины и их предметы сгруппированы по рубрикам Государственного рубрикатора научно-технической информации (ГРНТИ). Для определения содержания рубрик и отнесения к ним публикаций сформулированы связи терминов в виде логических условий, в которых определены предметные или проблемные границы. Разрабатываемая онтология положена нами в основу систематизации и интеграции электронных научных информационных 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 ГРНТИ с формированием соответствующего электронного каталога; 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 онтологий [3,4].

Демонстрационный прототип системы (содержащий, в частности, более подробное описание основных компонент ИАС и применяемых технологий) представлен в сети Интернет – <http://ias.krc.karelia.ru>. В полном объеме возможности ИАС (в первую очередь ГИС-сервисы) доступны пользователям корпоративной сети Интранет научных учреждений КарНЦ РАН.

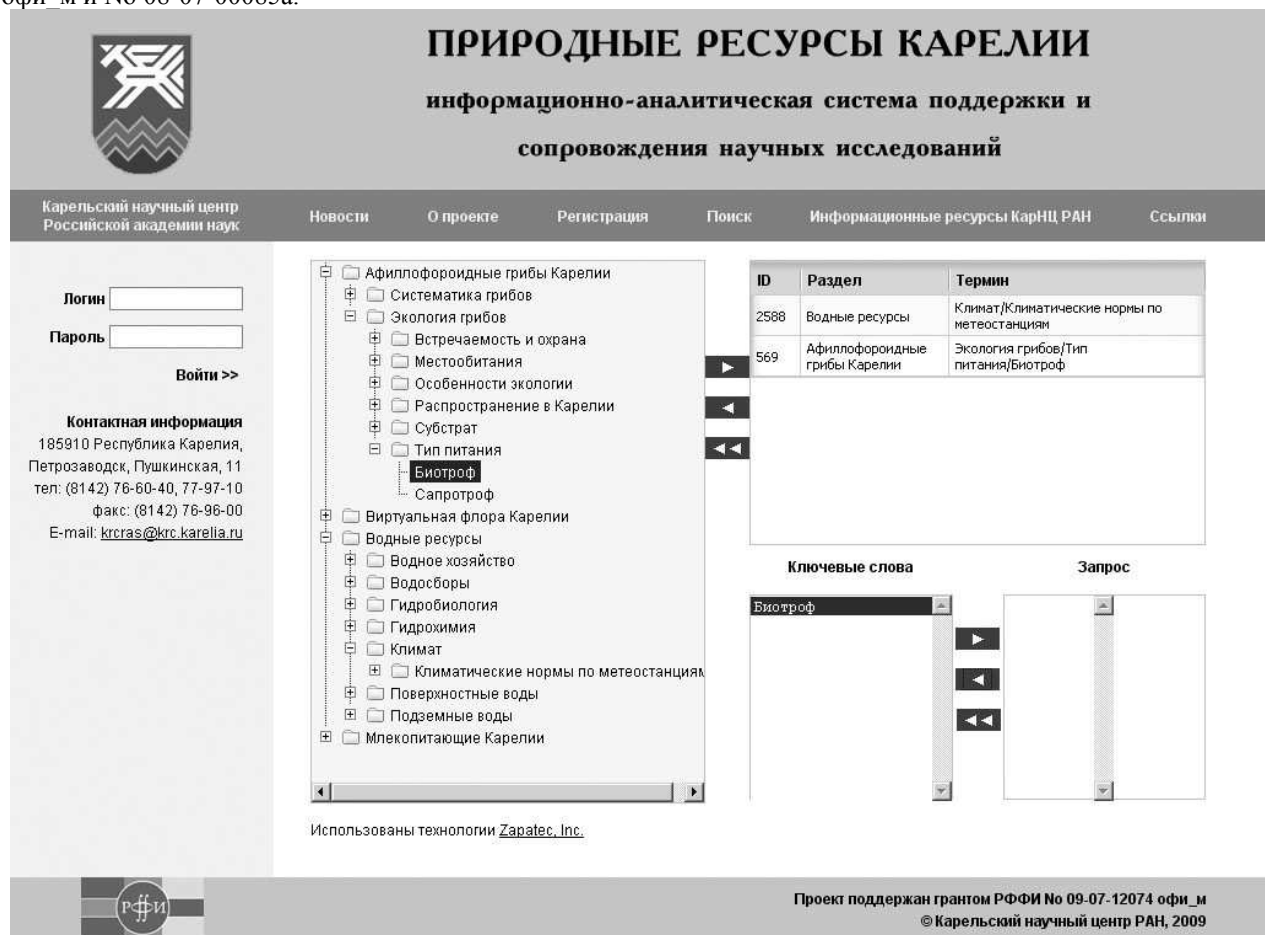


Рис. 1

ЛИТЕРАТУРА:

1. Добров Б.В., Лукашевич Н.В. и др. Разработка лингвистической онтологии по естественным наукам для решения задач информационного поиска //Труды VII Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции" – RCDL'2005, Ярославль, Россия, 2005.
2. Труды Симпозиума "Онтологическое моделирование". //Под ред. Л.А. Калиниченко М., ИПИ РАН, 2008 – 303 с.
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 VIII Всероссийской научной конференции "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 X Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Дубна, Россия, 7–11 октября 2008. С. 63–69.