

## ОРГАНИЗАЦИЯ И ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ РЕСУРСОВ

**ВДОВИЦЫН Владимир Трофимович** – кандидат физико-математических наук, доцент, заместитель председателя Карельского научного центра Российской академии наук (КарНЦ РАН) по научной работе, руководитель лаборатории информационных компьютерных технологий Института прикладных математических исследований (ИПМИ) КарНЦ РАН

*e-mail: vdov@krc.karelia.ru*

**ЛЕБЕДЕВ Виктор Алексеевич** – кандидат экономических наук, старший научный сотрудник лаборатории математической кибернетики ИПМИ) КарНЦ РАН

*e-mail: lebedev@krc.karelia.ru*

## **ТЕХНОЛОГИИ СИСТЕМАТИЗАЦИИ И ПОИСКА ЭЛЕКТРОННОЙ НАУЧНОЙ ИНФОРМАЦИИ С ПРИМЕНЕНИЕМ ОНТОЛОГИЙ**

### ***Введение***

Институты Карельского научного центра (КарНЦ) РАН проводят многолетние фундаментальные исследования: состояния и рационального использования природных (биологических, водных, лесных, минеральных) ресурсов (Институт биологии, Институт водных проблем Севера, Институт геологии и Институт леса); социальной и экономической инфраструктуры региона (Институт экономики); состояния, охраны и использования историко-культурного наследия населения Карелии (Институт языка, литературы и истории), методов математического моделирования и создание информационно-телекоммуникационных систем для поддержки и сопровождения научных исследований (Институт прикладных математических исследований).

Полученная в результате многолетних исследований уникальная научная информация в настоящее время частично уже представлена в виде цифровых коллекций научных информационных ресурсов, например, в электронной библиотеке – <http://dl.krc.karelia.ru>, локальных баз данных и отдельных Интернет- и ГИС-проектов (например, <http://lakemodel.net> – «ГИС – водные объекты Республики Карелия»). В 1997 году создан и функционирует официальный Web-сайт КарНЦ РАН (<http://www.krc.karelia.ru>), сайты институтов (<http://economy.krc.karelia.ru>, <http://illhportal.krc.karelia.ru> и др.), а также ряд тематических предметно-ориентированных сайтов с интегрированными базами данных, например, в области гуманитарных наук – <http://phonogr.krc.karelia.ru>, <http://toris.krc.karelia.ru> и др. Большинство этих проектов выполнялись в течение 1997–2010 гг. при финансовой поддержке Президиума РАН, грантов РФФИ, РГНФ и др. Для разработки информационных систем использовались традиционные Web-технологии и свободно распространяемое программное обеспечение (Linux, Apache, СУБД – PostgreSQL и MySQL) с организацией доступа к базам данных по разработанным web-интерфейсным формам с применением технологии CGI и PHP [1–4].

С целью развития, интеграции и эффективного использования электронных научных информационных ресурсов в настоящее время в КарНЦ РАН разрабатывается проект по созданию информационно-аналитической системы (ИАС) поддержки и сопровождения научных исследований «Природные ресурсы Карелии» – <http://ias.krc.karelia.ru>. Создание такой системы необходимо и важно, на наш взгляд, 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 (Инtranет/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 биоразнообразии и др. Представление этих научных материалов в виде единой (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, мониторингу изменения растительного покрова по данным дистанционного зондирования и т.п. Таким образом, на основе этой системы предполагается создать информационную основу для рационального использования природных ресурсов и развития региона.

### ***Систематизация и поиск электронной информации***

Эффективность поиска информации характеризуется тремя основными показателями: полнота, точность, затраты времени на поиск. Полнота поиска означает, что найдены все релевантные запросу публикации в заданном массиве. Однако сплошной просмотр всех публикаций в массиве приводит к существенному увеличению времени поиска. Для уменьшения последнего целесообразно систематизировать массив публикаций так, чтобы поиск выполнялся только в определенной части массива. Точность поиска (пертинентность) означает, что в отклике на запрос присутствуют именно те публикации, которые соответствуют информационным потребностям пользователя (и в идеальном случае отсутствуют публикации из других предметных областей). Одним из главных средств «борьбы» за точность поиска также является систематизация публикаций. Кроме того, повышению полноты и точности поиска способствует технология построения запросов, основанная на соответствующей систематизации предметных областей. Следует также отметить, что существенно сокращает время поиска индексация текстов публикаций. В этом случае вместо полнотекстового поиска по всему массиву публикаций выполняется поиск в базе данных индексов, что существенно быстрее.

Одним из самых распространенных средств систематизации публикаций являются различные рубрикаторы, в частности, Универсальный десятичный классификатор (УДК) и Государственный рубрикатор научно-технической информации (ГРНТИ).

В рамках создания ИАС «Природные ресурсы Карелии» для решения задач систематизации и организации поиска научной электронной информации нами разрабатывается подход, основанный на совместном применении ГРНТИ и методов онтологического моделирования.

Термин «онтология» ввел в оборот информатики американский ученый Т. Gruber (1993), заимствовав его из классической философии [5]. В самом общем виде онтология включает словарь терминов моделируемой предметной области и связей между ними. Вопросам построения и использования онтологий (общих, предметно-ориентированных) для разработки информационных систем нового поколения в настоящее время уделяется большое внимание. В качестве примера построения онтологии верхнего уровня следует отметить проект CYC (начат в 1994 г., Lenat D., <http://www.cyc.com>). Разработан ряд языков для описания онтологий – диалекты OWL, RDF/RDFS, Ontobroker и др. Наряду с созданием онтологий верхнего уровня, также большое значение (в т.ч. и для разработки информационных систем нового поколения) имеют специализированные предметные онтологии, такие как, например, предметная онтология в области медицины GALEN (<http://www.opengalen.org/>), культуры CIDOC CRM (<http://cidoc.ics.forth.gr/>) и др.

Различные аспекты применения методов онтологического моделирования в информационных системах (в т.ч. для повышения эффективности информационного

поиска) рассматривалось в докладах Симпозиума «Онтологическое моделирование» (Труды Симпозиума под ред. Л.А. Калиниченко, Звенигород, май 2008), а также в ряде других работ, например, [6–9]. Среди известных нам русскоязычных онтологий по естественным наукам, применяемых для индексации научных публикаций, необходимо отметить разрабатываемый в НИИ ВЦ МГУ под руководством Б.В. Доброва проект лингвистической онтологии [10].

### ***Разработка таксономий терминов***

В рамках создания ИАС «Природные ресурсы Карелии» мы предложили следующий подход к подбору и комплектации научных терминов в составе таксономии по направлениям естественных наук, исследования по которым проводятся в КарНЦ РАН.

В соответствии с ГРНТИ фиксируется состав научных дисциплин и предметов их изучения. В качестве предметов дисциплин могут выступать объекты, их свойства, строение, функционирование как в целом, так и отдельных частей и систем, взаимодействия с другими объектами, их роли и функции в составе систем, воздействие на них других компонент систем.

Объекты подразделяются на две категории: массовые и уникальные. Первые представляются видовыми именами по классификации в рамках суперкласса (организмы, минералы, почвы, рельеф и т.д.). Вторые представляются собственными именами (реки, озера, населенные пункты). В онтологию включаются названия суперклассов, классификации и собственные имена.

Свойства можно трактовать как отображение вида и степени отношения объекта с другими объектами, явлениями. При этом вид отношения – это название свойства, а степень выражается в виде некоторой шкалы. Для онтологии представляют интерес шкалы значений свойств, выражаемых соответствующими терминами. В этом случае свойство представляет собой классификацию (например, показатели трофности озер: дистрофное, олиготрофное, мезотрофное, эвтрофное).

Строение описывается морфологией и анатомией. В онтологию включаются названия частей, органов, систем и их классификации, как и в случае объектов.

Функционирование – это процесс, в котором части объекта взаимодействуют, выполняя определенные функции. Для различных суперклассов объектов под функционированием понимают свои процессы: например, в биологии – это физиология, сукцессии; в водных объектах – течения, русловые процессы; в геологии – тектоника, выветривание. В онтологии процесс представляется набором терминов, обозначающих этапы (фазы, стадии) процесса, взаимодействующие части (объекты), реакции, преобразования, их результаты. Например, в физиологии растений одним из этапов является фотосинтез, который состоит из световой и темновой стадий. Световая стадия протекает в системе, содержащей светособирающие компоненты, электротранспортные цепи, хлоропласты, хлорофиллы, механизм Ферстера. Процесс состоит в усвоении солнечной энергии путем ее различных преобразований и синтеза конечного носителя энергии АТФ (аденозинтрифосфата). В онтологию должны быть включены все термины промежуточных преобразований света, реакции и их продукты. Аналогично описывается темновая стадия, дыхание и т.д. Подробность описания определяется степенью изученности процесса.

Таким образом, таксономия онтологии по естественным наукам в нашем случае представляет собой иерархическую структуру, основанием которой является выборка рубрик ГРНТИ, отображающих направления научных исследований институтов КарНЦ РАН. Эти рубрики охватывают соответствующие научные дисциплины и предметы их изучения. На более высоких уровнях таксономии представлены: классификационные схемы объектов и систем, относящихся к предметам дисциплин; классификаторы свойств объектов и их значений, отображающих участие объектов в системе более высокого уровня (в том числе возможное отношение к экономике); составные части объектов и их

классификации и свойства, имеющие важное значение для функционирования объектов. На **рис.1–2** показаны фрагменты ГРНТИ и таксономии по экологии растений (рубрика «Организм и среда»).

#### Поиск по рубрикатору ГРНТИ

- выделить все    убрать выделение
- 34. Биология
    - 34.01. Общие вопросы
      - 34.01.21. Организация научно-исследовательских работ
    - 34.19. Цитология
      - 34.19.17. Цитоморфология
      - 34.19.19. Цитофизиология
      - 34.19.23. Цитозология. Реакция клетки на внешнее воздействие.
    - 34.23. Генетика
      - 34.23.57. Генетика возделываемых растений и лесных пород
      - 34.23.59. Генетика домашних и сельскохозяйственных животных
    - 34.29. Ботаника
      - 34.29.01. Общие вопросы
      - 34.29.15. Споровые растения
      - 34.29.25. Семенные растения
      - 34.29.35. Растительность. Фитоценология.
    - 34.31. Физиология растений
      - 34.31.01. Общие вопросы
      - 34.31.15. Действие физических факторов на растения.
      - 34.31.19. Дыхание и обмен веществ в растениях
      - 34.31.27. Рост и развитие растений
      - 34.31.37. Биология и физиология культурных растений
    - 34.33. Зоология
      - 34.33.15. Зоология беспозвоночных
      - 34.33.19. Энтомология
      - 34.33.23. Зоопаразитология
      - 34.33.27. Зоология наземных позвоночных
      - 34.33.33. Ихтиология

Рис. 1. Фрагмент ГРНТИ

#### Поиск по онтологии

- выделить все    убрать выделение
- Афиллофоридные грибы Карелии
    - Систематика грибов
      - Систематика грибов (лат)
      - Систематика грибов (рус)
    - Экология грибов
      - Встречаемость и охрана
      - Местобитания
      - Особенности экологии
      - Распространение в Карелии
      - Субстрат
      - Тип питания
  - Виртуальная флора Карелии
    - Систематика растений
      - Систематика растений (лат)
      - Систематика растений (рус)
    - Экология растений
      - Жизненная форма
      - Роль в растительном покрове
      - Хозяйственное значение
      - Экологические характеристики
      - Экотип
      - Распространение в Карелии
      - Встречаемость и охрана
      - Местобитания
  - Млекопитающие Карелии
    - Систематика млекопитающих
      - Систематика млекопитающих (лат)
      - Систематика млекопитающих (рус)
    - Экология млекопитающих

Рис. 2. Фрагмент таксономии

Термины в таксономии связаны типами отношений классификации (род-вид), агрегации (часть-целое) и синонимии. Два первых типа формируют иерархическую структуру таксономии. Синонимия терминов образует отдельную таблицу, записи которой - синонимические гнезда по отношению к доминантному термину.

Помимо указанных типов отношений между терминами существуют слова, входящие в состав многословных терминов (устойчивые словосочетания, синтагматические связи), например, леса, луга, длина, биомасса. Значения их определяются именно в составе словосочетаний, что порождает определенные трудности их использования в таксономии терминов и при поиске. Для разрешения такого рода коллизий разработан ряд приемов, которые описаны в работе [11].

В настоящее время разработан значительный фрагмент таксономии терминов по естественным наукам, а также технологии ее загрузки, редактирования и использования для индексации публикаций и информационного поиска.

#### ***Систематизация электронных публикаций***

Систематизация публикаций предназначена для их разделения по темам с целью сокращения времени поиска по запросам и выполняется с использованием онтологии. Предполагается, что массивы публикаций сопровождаются метаданными, в состав которых обязательно включаются заголовки публикаций и списки ключевых слов. Процесс систематизации с целью сокращения затрат времени на его выполнение разделяется на два этапа: предметизацию и индексацию. Необходимо заметить, что технологии предметизации хорошо разработаны и достаточно давно используются в библиотечном деле, например, [12].

В нашем случае предметизация публикаций осуществляется на основе анализа их названий и ключевых слов. В результате выполняется отнесение публикаций к соответствующим рубрикам ГРНТИ. Затем, на этапе индексации, с целью выявления имеющейся в их составе терминологии текст публикаций сопоставляется с фрагментом таксономии онтологии, отнесенном к соответствующей рубрике.

Технология индексации опубликована в ряде наших работ, например, в [13]. Она состоит в «поэлементном» просмотре программой индексации текстов публикаций, выявления терминов онтологии, определении их мест в составе онтологии и выделении соответствующих фрагментов онтологии, используемых в качестве индекса публикации. Результатом индексации массива публикаций является база их индексов, которая впоследствии используется для поиска по запросам.

В качестве информационной основы предметизации (кроме таксономии) используется набор логических условий отнесения публикаций к соответствующим рубрикам ГРНТИ. Для формирования этих условий используется ряд номенклатур из таксономии терминов. Для разработки условий соответствия публикаций рубрикам ГРНТИ необходимо было проанализировать достаточно представительный массив научных публикаций сотрудников КарНЦ РАН. Его мы нашли в составе электронной библиотеки КарНЦ РАН (dl.krc.karelia.ru). В результате проведенного анализа были определены следующие типы публикаций по характеру работ относительно к ГРНТИ:

- описание результатов экспериментов, наблюдений, мониторинга и технологий;
- обобщенное описание объектов исследований, разработок;
- состояние, проблемы и перспективы научных дисциплин, междисциплинарных исследований (общие вопросы по дисциплинам, наукам).

Для каждого типа публикаций разработана обобщенная схема логического условия:

- <объект эксперимента> **AND** (<объект его целое> **OR** <объект его часть> **OR** <действующий фактор> **OR** <действие> **OR** <результат>) **AND** <границы, ограничения>;
- <объект описания> **AND** (<тема> **OR** <пусто>);
- <дисциплина> **AND** (<характеристика> **OR** <пусто>).

Здесь термины в угловых скобках (нетерминалы) символически представляют номенклатуры терминов, являющихся частью таксономии терминов. Для каждой рубрики ГРНТИ это будут свои номенклатуры (хотя одна и та же номенклатура может входить в условия разных рубрик), **AND** и **OR** и **NOT** – логические операции конъюнкции, дизъюнкции и отрицания.

Далее представлены примеры логических условий для некоторых рубрик ГРНТИ. Следует отметить, что логические условия формируются для каждой рубрики индивидуально с использованием указанных схем. При этом в зависимости от содержания рубрики, определяемого экспертно, логическое условие может состояться как комбинация из указанных схем.

#### 34.29.35. Растительность. Фитоценология.

фитогеография **OR** фитоценология **OR** ((растительность **OR** сообщество **OR** фитоценоз) **AND** (лес **OR** луг **OR** болота **OR** вырубки **OR** окрестность **OR** тайга **OR** водосбор **OR** остров **OR** приморский **OR** пригород **OR** динамика **OR** развитие))

#### 37.27.27. Физические свойства вод суши. Гидрохимия.

гидрохимия **OR** <список классов объектов вод суши> **OR** донные отложения **AND** (химия **OR** физика **OR** оптика **OR** плотность **OR** прозрачность **OR** цвет **OR** мутность **OR** электропроводность **OR** <газ>

**OR** ион **OR** <катион> **OR** <анион> **OR** <биоген> **OR** закисление **OR** загрязнение **OR** <элемент>) **AND NOT** (море **OR** морской)

Названия в угловых скобках представляют соответствующую номенклатуру или классификацию. Для наглядности некоторые номенклатуры показаны в условии списками компонент. В настоящее время сформулирован ряд условий предметизации публикаций по биологии, почвоведению, лесному хозяйству и водным ресурсам, относящихся к научным направлениям КарНЦ РАН.

В соответствии с процедурой предметизации сначала определяется формат метаданных коллекции, после чего читается заголовок первой публикации и список ключевых слов (если он имеется), поочередно в этом тексте слова выделяются и сопоставляются с терминами логического условия первой по порядку рубрики ГРНТИ. В результате условие может принять значение «истина» или «ложь». В первом случае в базу индексов записывается адрес публикации и номер рубрики. Далее производится сопоставление текста (заголовка и списка ключевых слов) с условием предметизации следующей рубрики. Эта процедура выполняется для всего заданного списка рубрик. При этом одна и та же публикация может быть предметизирована на одну, две и более рубрик ГРНТИ.

По аналогичной схеме выполняется предметизация по всем имеющимся в информационной системе коллекциям публикаций, в результате формируется база индексов, необходимая для выполнения индексации публикации, во время которой по номерам рубрик выделяются соответствующие фрагменты таксономии терминов и формируются индексы также в виде фрагментов таксономии. При этом необходимо отметить, что метаданные публикаций коллекции помечаются, чтобы не предметизировать их повторно после очередного пополнения коллекции новыми публикациями.

Публикации, для которых попытка предметизации не дала результата, подвергаются повторной предметизации при каждом запуске алгоритма в расчете на возможное пополнение набора условий предметизации.

Для иллюстрации работы алгоритма предметизации на **рис. 3** представлена расшифровка проверки выполнения условия для одной из публикаций.

**Проблемы изучения, рационального использования и охраны ресурсов Белого моря**

Условие отнесения публикации к рубрике ГРНТИ 34.35.33. Гидробиология:

```
((реки|водоем|озера|море|водоем|водохранилища|вода|водный объект)
&& (экология|биоразнообразие|флора|фауна|трава|планктон|планктер|перифитон|
обрататель|популяция|поселение|биоценоз|гидробиоценоз|бентос|биомасса|
круговорот веществ|ресурс|водоросль|биота|беспозвоночные|сообщества|продукт)
&& ((остров|болота|геолог)))
|| ((реки|водоем|озера|море|водоем|водохранилища|вода|водный объект)
&& (экосистема|биоценоз|сообщество)
&& (состояние|охрана|устойчивость|нарушенность|антропогенные воздействия)
&& ((остров|болота|геолог)))
|| (аквакультура|амфипод|мидии|бокоплав|фукус|ламинария|Amphipod|Crustace|Mytilus|
Littorin|Nereis|Mya|перифитон|макрофит)
```

Статья относится к разделу 34.35.33 - **Гидробиология** (ключевые слова: моря, ресурсов, охраны)

Условие отнесения публикации к рубрике ГРНТИ 37.25.33. Региональная океанология:

```
(Белое море|морская вода)
&& ((побережье|остров|берег|бассейн|водосбор|лес|болота|реки |приток|массив|месторождение|
экосистема|биоценоз|биоразнообразие|биоресурс|беспозвоночные|бентос))
&& (отложения|ресурс|гидрофизический режим|динамика вод|гидрохимия|химия вод|
гидродинамика|течение|приливы|термика|ледовый режим|ледоставный период|опреснение|
уровень|температура|литораль|пелагиаль|исследование|изучение|биоген|под влиянием)
```

Статья относится к разделу 37.25.33 - **Региональная океанология** (ключевые слова: белого, моря, изучения, ресурсов)

Рис. 3. Расшифровка проверки выполнения условия для одной из публикаций

### **Тематический поиск**

Разработанные таксономия терминов и база индексов публикаций обеспечивают тематический поиск публикаций в их массивах по запросам пользователей. Простейший

вид запроса состоит в требовании показать все публикации, относящиеся к тематике одной из рубрик ГРНТИ. Пользователю предлагается сделать выбор рубрики по рубрикатору. После чего в базе индексов находятся записи, содержащие ее номер, и список названий публикаций визуализируется на экран в виде гиперссылок для последующего просмотра или печати текстов публикаций.

В общем случае тематический запрос на поиск релевантных публикаций может быть достаточно сложным, например, «можно ли найти жаропонижающее лекарственное растение на сухой опушке смешанного леса». Для обеспечения построения «правильных» запросов и сокращения времени поиска разработана технология построения запросов на основе таксономии терминов.

Пользователю сначала предлагается выбрать рубрику ГРНТИ, которая, по его мнению, должна содержать материалы по его запросу (если этих рубрик не одна, то придется построить несколько однотипных запросов). Тогда ему предлагается соответствующий рубрике фрагмент таксономии, в котором он должен отметить интересующие его термины. С использованием этих терминов формулируется запрос в виде логического выражения, конъюнктивные и дизъюнктивные связи терминов. Вид выражения демонстрируется пользователю на экране, чтобы он мог его оценить и исправить в случае необходимости.

При выполнении запроса каждый термин трактуется как левая часть выражения  $T=X$ , где  $T$  – термин запроса,  $X$  – термин в базе индексов. Все термины индекса сравниваются с терминами запроса и, если логическое выражение запроса принимает значение «истина», то название соответствующей публикации помещается в отклик в виде гиперссылки на публикацию.

На **рис.4** показаны этапы выбора терминов из таксономии и формирования запроса в ИАС «Природные ресурсы Карелии».

**ПРИРОДНЫЕ РЕСУРСЫ КАРЕЛИИ**  
информационно-аналитическая система поддержки и сопровождения научных исследований

Карельский научный центр Российской академии наук

Новости О проекте Регистрация Поиск Информационные ресурсы КарНЦ РАН Ссылки

Личные данные  
Собранные результаты  
Выйти из кабинета

Контактная информация  
185910 Республика Карелия,  
Петрозаводск, Пушкинская, 11  
тел: (8142) 76-60-40, 77-97-10  
факс: (8142) 76-96-00  
E-mail: [krccas@krcc.karelia.ru](mailto:krccas@krcc.karelia.ru)

Афиллофоридные грибы Карелии  
 - Систематика грибов  
 - Экология грибов  
 - Встречаемость и охрана  
 - Местообитания  
 - Особенности экологии  
 - Распространение в Карелии  
 - Субстрат  
 - Тип питания  
 - Биотроф  
 - Сапротроф

ID	Раздел	Термин
2595	Водные ресурсы	Гидробиология
569	Афиллофоридные грибы Карелии	Экология грибов/Тип питания/Биотроф

Ключевые слова: Биотроф, Гидробиология  
Запрос: Гидробиология

Для ключевых слов: (Гидробиология) найдено 4 документа

Документы коллекции "Электронные версии научных публикаций, изданных в РИО КарНЦ РАН"

- Библиография работ по водорослям Европейского Севера России (Республика Карелия, Мурманская область)
- Биологические ресурсы Белого моря и внутренних водоемов Европейского Севера. Материалы XXVIII Международной конференции 5–8 октября 2009 г.
- Биотрансформация и круговорот органических веществ в экосистеме Белого моря: оценка на основе математического моделирования
- К оценке продукционных возможностей папариги Двинского залива Белого моря

Записать результаты в портфель | выделить все | убрать выделение

Использованы технологии Zscaler, Inc.

Проект поддержан грантом РФФИ № 09-07-12074 офи\_м  
© Карельский научный центр РАН, 2009-2010

Рис. 4. Этапы выбора терминов из таксономии и формирования запроса

### **Заключение**

В результате работы по проекту создания ИАС «Природные ресурсы Карелии» сформирована и развивается интегрированная информационная база научной информации по природным ресурсам региона, включающая картографические материалы, атрибутивные базы данных, электронные научные публикации и коллекции. В качестве основы интеграции разнородной научной информации применена комплексная онтология по естественным наукам, которая разработана совместно со специалистами-предметниками из научных учреждений (биологии, водных проблем Севера, геологии, леса и прикладных математических исследований) КарНЦ РАН и охватывает предметные области исследований участников проекта. Она является основой разработанных технологий систематизации контента и поиска, позволяющих пользователю получать по запросу комплекс взаимосвязанной научной информации (тематические слои, соответствующую атрибутивную информацию из баз данных, а также электронные публикации, документы из электронных библиотек и т.п.), релевантный его информационным потребностям.

Работы по созданию ИАС «Природные ресурсы Карелии» 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.В., Старковой В.Г., а также всем сотрудникам научных учреждений КарНЦ РАН, участвующих в формировании комплексной онтологии, соответствующих тематических слоев и атрибутивной информации.

### **Литература:**

1. Вдовицын В., Сорокин А. Вопросы формирования и использования электронных научных информационных ресурсов // Информационные ресурсы России. - 2004. - №4. - С. 7–12.
2. Вдовицын В.Т., Сорокин А.Д., Луговая Н.Б. Электронная библиотека научных информационных ресурсов КарНЦ РАН: состояние и перспективы развития // Труды Шестой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2004, Пуцино, 29 сентября – 1 октября 2004 г. - С. 41–46.
3. Керт Г.М., Вдовицын В.Т. Информационные технологии в исследовании топонимии // Вопросы языкознания. - 2005. - № 3. - С. 102–124.
4. Бархатов А.В., Вдовицын В.Т. и др. Электронные научные информационные ресурсы для поддержки инвестиционной деятельности в регионе // Информационные ресурсы России. - 2006. - № 4(92). - С. 14–17.
5. Gruber T. R. A Translation Approach to Portable Ontology Specifications // Knowledge Acquisition Journal. - 1993. - Vol.5. - P. 199–220.
6. Рогушина Ю.В. Использование онтологического описания предметной области для повышения релевантности информационного поиска // Проблемы программирования. – 2003. – № 4. – С. 54–64.
7. Гладун А.Я., Рогушина Ю.В. Применение тезауруса предметной области для повышения релевантности поиска в Интернете // Искусственный интеллект. - 2005. - №4. - С.742–752. [Электронный ресурс]. - Режим доступа: [www.iai.dn.ua/public/JournalAI\\_2005\\_4/Razdel8/02\\_Gladun\\_Rogushina.pdf](http://www.iai.dn.ua/public/JournalAI_2005_4/Razdel8/02_Gladun_Rogushina.pdf).



8. Загорюлько Ю.А. Методы и методологии разработки, сопровождения и реинжиниринга онтологий // Труды Симпозиума «Онтологическое моделирование» под ред. Л.А. Калиниченко, Звенигород, 2008. - С. 167–200.
9. Россеева О.И., Загорюлько Ю.А. Организация эффективного поиска на основе онтологий. [Электронный ресурс]. - Режим доступа: [http://www.dialog-21.ru/Archive/2001/volume2/2\\_49.htm](http://www.dialog-21.ru/Archive/2001/volume2/2_49.htm).
10. Добров Б.В., Лукашевич Н.В. и др. Разработка лингвистической онтологии по естественным наукам для решения задач информационного поиска // Труды Седьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2005, Ярославль, Россия, 2005.
11. Лебедев В.А. Роли онтологий в электронной библиотеке КарНЦ РАН // Труды XI Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2009. Петрозаводск: КарНЦ РАН, 2009. - С. 370–375.
12. Руководство по методике предметизации. Опыт Российской национальной библиотеки, 2005. [Электронный ресурс]. - Режим доступа: <http://www.nlr.ru/cat/predmet/method/index.html>.
13. 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» – RCDL'2008 (Дубна, 7–11 октября 2008 г.). Дубна: ОИЯИ, 2008. - С. 63–69.