

Технологии информационного обеспечения научных исследований в ИАС «Природные ресурсы Карелии»

Введение

В настоящее время проблема рационального использования огромных хранилищ электронной информации в интересах науки, образования, разработки и развития высокотехнологичных производств остается актуальной. Одним из подходов к решению этой проблемы является создание распределенных компьютерных систем для информационной поддержки междисциплинарных исследований с использованием Web, ГИС-технологий и методов онтологического моделирования [1-4]. При построении такого рода систем необходимо разрабатывать и применять эффективные технологии систематизации и поиска электронной научной информации, которые позволяли бы специалистам получать релевантный запросу комплекс взаимосвязанной научной информации (электронные научные публикации, тематические коллекции документов, базы данных, ГИС-системы и т.п.), а также предусмотреть автоматическое пополнение хранилищ информационной системы новой научной информацией. В такой системе необходимо обеспечить для пользователя возможность получать по запросу (например, путем выделения определенного участка карты) соответствующую научную информацию (электронные публикации, базы данных, документы коллекций и т.п.), которая семантически связана с выбранным участком электронной карты.

Традиционные методы информационного поиска, основанные на использовании ключевых слов, обладают рядом недостатков, связанных, например, с многозначностью (polysemous) используемых в запросе терминов, а также недостаточным знанием пользователями терминологии самой предметной области. Одним из перспективных направлений исследований и разработок в плане повышения эффективности информационного поиска является применение методов онтологического моделирования (ontology-based information retrieval) [5-10]. Такие системы информационного поиска учитывают смысловое содержание терминов запроса, используют онтологии как для индексации информационных ресурсов, так и для организации семантического поиска.

В данной статье представлена общая архитектура разрабатываемой нами информационно-аналитической системы (ИАС) для поддержки научных исследований, а также разработанные и используемые при построении ИАС технологии систематизации и поиска научной информации с применением онтологий. Предложен подход к решению задачи географического поиска электронных научных публикаций.

Архитектура ИАС

«Природные ресурсы Карелии»

Создание информационно-аналитической системы для поддержки научной, аналитической и управленческой 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. Для достижения поставленной цели, на наш взгляд, следует в первую очередь обеспечить автоматизированный сбор, систематизацию и эффективный доступ ученых и специалистов к необходимой научной информации. Исходя из решения этих задач, архитектура разрабатываемой нами системы в общем виде выглядит следующим образом (рис. 1).

Основные компоненты системы можно кратко описать следующим образом:

- **Сервис поиска тематической электронной научной информации в сети Интернет.** Этот сервис основан на применении тематического Web-краулера [18], который в процессе своей работы формирует в ИАС временное хранилище электронных научных публикаций, карт, космических снимков и соответствующих метаданных, полученных из сети Интернет в результате тематического поиска.
- **Сервис систематизации электронной научной информации.** Осуществляет автоматическую систематизацию (предметизацию и индексацию) электронных научных публикаций с использованием онтологии, а также

ВДОВИЦЫН Владимир Трофимович - кандидат физико-математических наук, доцент, заместитель председателя КарНЦ РАН по научной работе, руководитель лаборатории информационных компьютерных технологий Института прикладных математических исследований Карельского научного центра Российской академии наук (ИПМИ КарНЦ РАН).
Адрес: Республика Карелия, 185610, г. Петрозаводск, ул. Пушкинская, 11
e-mail: vdov@krc.karelia.ru

ЛЕБЕДЕВ Виктор Алексеевич - кандидат экономических наук, научный сотрудник лаборатории информационных компьютерных технологий ИПМИ КарНЦ РАН.
Адрес: Республика Карелия, 185610, г. Петрозаводск, ул. Пушкинская, 11
e-mail: lebedev@krc.karelia.ru

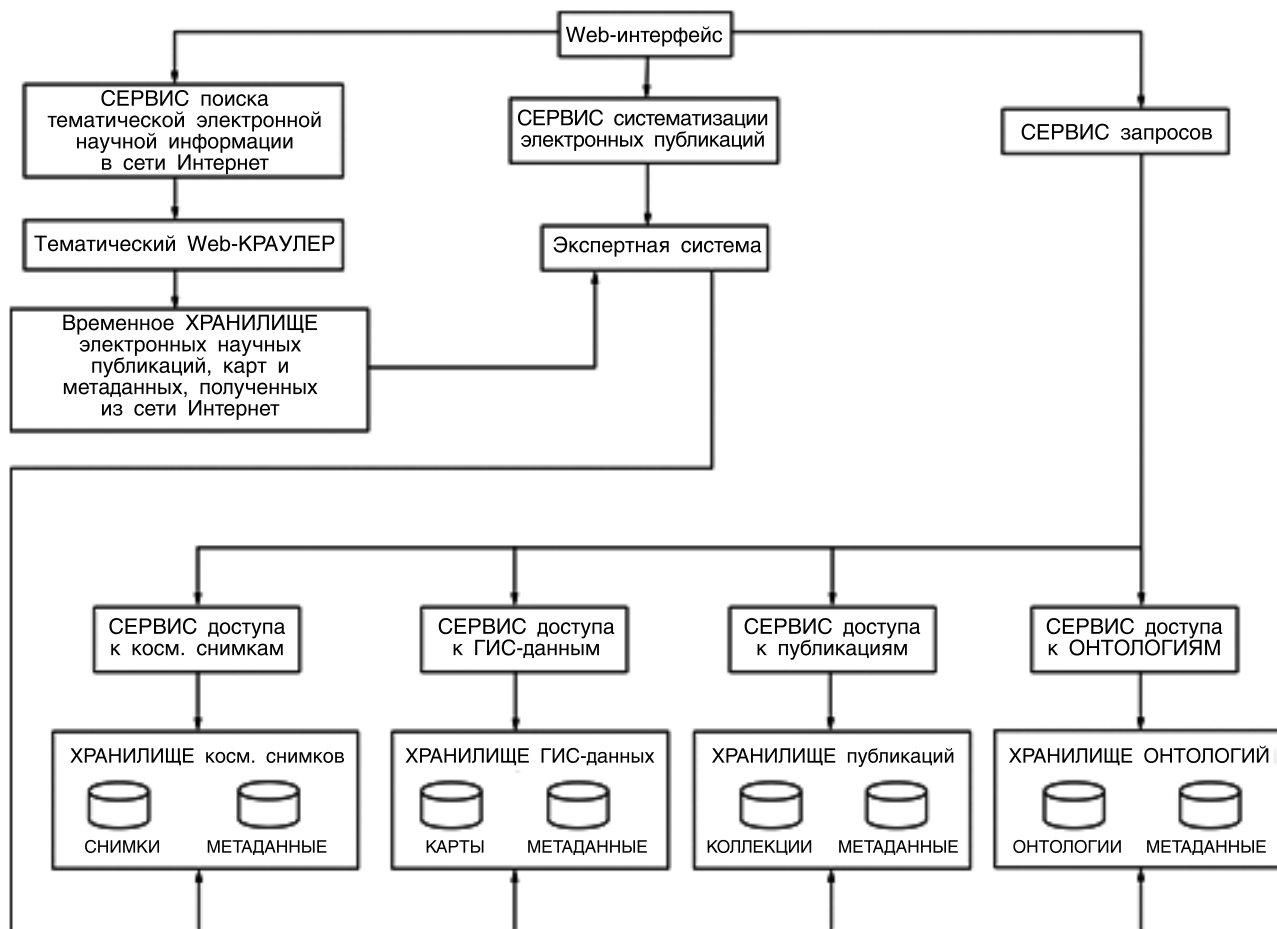


Рис. 1. Архитектура ИАС «Природные ресурсы Карелии» - основные компоненты

пополнение и корректировку хранилищ системы: космических снимков, ГИС-данных, электронных научных публикаций и онтологий. Процедура систематизации электронных научных информационных ресурсов основана на применении технологии продукционных экспертных систем (ЭС), которая содержит набор правил-продукций (условие - логическое выражение (описывает содержание публикации), а действие - отнесение публикации к определенной рубрике (в нашем случае - к определенным рубрикам ГРНТИ)). На входе ЭС - массив электронных публикаций (отобранный из временного хранилища системы); на выходе - распределение электронных публикаций по рубрикам ГРНТИ (предметизация). Далее на этапе индексации с помощью онтологии автоматически формируется база индексов электронных публикаций.

- **Сервис запросов.** Осуществляет поддержку пользователей при составлении запросов к ИАС; выполнение процедуры поиска в хранилищах системы с использованием автоматически сформированного системой логического условия отбора данных и базы индексов публикаций; ранжирование результатов поиска по степени релевантности запросу в соответствии со специальным алгоритмом; отбор и сохранение полученных результатов в «личном кабинете» пользователя.

Таким образом, разрабатываемая нами система должна обеспечить автоматизированный сбор тематической научной информации в сети Интернет, ее систематизацию (т.е. автоматическое разнесение электронных публикаций по их содержанию к определенным предметным рубрикам и формирование индексов), а также эффективный доступ пользователей к необходимой информации по запросам.

Обзор существующих подходов к построению информационных систем, базирующихся на применении методов онтологического моделирования

Исследования и разработки в области создания эффективных методов и технологий систематизации и поиска электронной информации в интернете на основе онтологического подхода проводятся как у нас в стране, так и за рубежом. В этой связи, на наш взгляд, необходимо отметить ряд следующих работ.

В работе [3] исследуются вопросы применения онтологического подхода для решения проблемы оперативного доступа к распределенным и разнородным источникам данных (фотографии биологических видов и среды их обитания, карты, электронные публикации, метаданные и т.п.) применительно к развитию сервисов ин-

формационной системы поддержки исследований в области биоразнообразия. В статье представлено описание Web-сервиса, который поддерживает обмен онтологиями и управление ими по сети. При этом разработан широкий спектр операций по созданию, хранению, управлению, анализу и интеграции различных онтологий данной предметной области.

В статье [5] предлагается использовать знания пользователя об интересующей его Предметной области (ПрО), представленные в виде онтологии. На основе множества терминов онтологии ПрО строится тезаурус пользователя, который используется для оценки того, насколько интересен этот информационный ресурс (ИР) пользователю. Предлагается алгоритм определения оценки соответствия ИР информационным потребностям пользователя, который реализован на языке программирования Java (OntologySearch). При этом разработка, редактирование и выбор терминов для формулирования запроса с помощью онтологии выполняется с использованием возможностей редактора онтологий Protégé. Все это, по мнению авторов, должно способствовать повышению релевантности запросов пользователя к информационной системе.

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 [6].

В работе [7] предлагается использовать онтологии для описания семантики представленных в сети ресурсов. Это позволит, по мнению авторов, решить проблему «несовместимости» используемых в запросах пользователя терминов (ключевых слов). При этом специально разработанные онтологии предполагается применять в качестве посредников между пользователем и информационной системой. За счет этого можно осуществить соответствующее расширение условия поиска, в частности, решить проблему синонимии терминов.

В работе [8] рассмотрены проблемы и методы информационного поиска. Предлагается метапоисковая система с интерфейсом в виде виртуального каталога, работа которой основана на применении онтологий предметных областей. При этом для поиска информации в интернете применены иерархия онтологий предметных областей, онтология Интернет-ресурсов и онтология пользователя.

Работа [9] посвящена проблеме поиска информации в web с помощью поисковых машин (web search engines). При этом отмечается, что основные трудности при поиске связаны с многозначностью (polysemous) ключевых слов, используемых пользователем в запросе. В статье предлагаются разработанные авторами методы семантического поиска, основанные на применении онтологий, позволяющие динамически определять возможные категории, к которым должна относиться найденная по запросу информация.

Также следует отметить работы [4, 10] в которых разрабатываются схожие с нашим подходы к решению проблемы систематизации и поиска электронной научной информации на основе применения методов онтологического моделирования.

Технологии систематизации и поиска информации с применением онтологий

В рамках создания и развития ИАС «Природные ресурсы Карелии» для решения задач систематизации и поиска научной электронной информации нами разрабатывается подход, основанный на совместном применении ГРНТИ и методов онтологического моделирования [11-16].

Систематизация электронных публикаций необходима 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 разделяется на два этапа: предметизацию и индексацию. При этом в качестве информационной основы предметизации (кроме таксономии терминов) используется набор логических условий, с помощью которых осуществляется процесс отнесения публикаций к соответствующим рубрикам ГРНТИ. Для формирования этих условий, описывающих содержание публикаций, используется ряд номенклатур из таксономии терминов.

Ниже приведен пример логического условия предметизации, представленного в виде правила-продукции ЭС.

IF (фитогеография OR фитоценология OR геоботаника OR растительность OR сообщество OR фитоценоз OR ценофлора);

THEN рубрика ГРНТИ - 34.29.35. Растительность. Фитоценологии.

В настоящее время сформулирован ряд логических условий для предметизации публикаций по биологии, почвоведению, лесному хозяйству и водным ресурсам, относящихся к научным направлениям КарНЦ РАН.

В процессе индексации в тексте каждой публикации ищутся термины соответствующего рубрике фрагмента таксономии и определяется их место в таксономии. Общее правило: если находится пара терминов, лежащих на одной ветви таксономии (за исключением пар, содержащих название рубрики), то эта ветвь помещается в индекс. Имеются исключения, например, ветви таксономии, содержащие систематику растений (или животных), помещаются в индекс, если в тексте статьи содержится хотя бы один термин систематики растений (или животных). Если ветвь таксономии является «малоуровневой» (содержит только разделы рубрикатора ГРНТИ - три уровня - и еще один уровень таксономии), то для помещения такого рода ветви в индекс требуется наличие в ней тоже только одного термина из текста статьи, не совпадающего с названиями рубрик. В результате индекс (будем называть его «содержательным индексом»

электронной публикации) представляет собой ряд строк (ветвей таксономии), начиная с названия рубрики (корня) и включая все термины таксономии вплоть до найденного термина.

Таксономия терминов и база индексов публикаций обеспечивают тематический поиск публикаций по запросам пользователей. Нами разработана технология построения запросов с использованием таксономии терминов, суть которой заключается в следующем. 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). Далее ему предлагается соответствующий рубрике фрагмент таксономии, в котором он должен отметить интересующие его термины (рис. 2). С использованием указанных терминов система автоматически формирует запрос в виде логического выражения, определяющего конъюнктивные и дизъюнктивные связи терминов. Необходимо отметить, что в первой версии сервиса поиска пользователю предлагалось самому составить конъюнктивно-дизъюнктивное логическое выражение отбора данных из выбранных по таксономии терминов [15]. Практика применения такого подхода пользователями при построении запросов в ИАС выявила ряд трудностей, связанных с процедурой составления условия отбора (например, во многих случаях пользователь формировал «бессмысленные» логические комбинации терминов). В связи с этим сервис поиска был модернизирован в плане автоматизации процесса построения условия отбора данных.

Следует отметить, что поскольку поиск по запросу осуществляется в базе индексов (а не в текстах электронных публикаций), запрос автоматически расширяется включением в него конъюнкции терминов от корня

и дизъюнкции терминов и их синонимов вплоть до листьев от указанных пользователем терминов. Тем самым обеспечивается повышение точности отклика на запрос за счет конъюнкции терминов предыдущих уровней таксономии и полноты за счет дизъюнкции терминов нисходящих уровней таксономии и их синонимов. Список названий публикаций визуализируется на экран в виде гиперссылок для последующего просмотра или сохранения текстов публикаций в «личном» кабинете пользователя.

В настоящее время ранжирование документов в отклике на запрос выполняется по следующим правилам. Первый ранг назначается документам, в которых полный набор терминов запроса встречается в его заголовке и аннотации. Далее определяется встречаемость набора терминов запроса в тексте документов и вычисляется отношение этого числа к числу страниц текста. Если это отношение не меньше половины, то документу присваивается второй ранг, а если это отношение меньше 0,5 - третий ранг. После чего выполняется упорядочивание документов отклика в соответствии с назначенными рангами.

Для оценки эффективности предложенных технологий систематизации и поиска электронных публикаций в ИАС были проведены вычислительные эксперименты с целью оценки полноты, релевантности и пертинентности запросов и сравнение с поисковой системой Яндекс (программа «Персональный поиск Яндекса», версия 2.6.0.). Анализ полученных результатов показал, что эффективность поиска в ИАС с использованием онтологии существенно выше, чем аналогичный поиск по Яндексу. В среднем эффективность поиска с использованием онтологии, по нашим оценкам, выше: по полноте - в 1,8 раза, а по точности и пертинентности - в 1,4 раза [16].

Поиск по таксономии терминов

выделить все | убрать выделение

- [-] 34. Биология (1152)
 - [-] 34.29. Ботаника (659)
 - [-] 34.29.25. Семенные растения (119)
 - [-] Сосудистые семенные растения Карелии (119)
 - [-] Экология растений (108)
 - [-] Роль в растительном покрове (3)
 - Дернообразующий (0)
 - Кочкообразующий (0)
 - Лесообразующий (2)
 - Обогащающее почву азотом (0)
 - Пионерный (1)
 - Подлесок (0)
 - Подстилкообразующий (0)
 - Сорное (3)
 - Торфообразующий (0)
 - [-] 34.35. Экология (420)
 - [-] 34.35.25. Биоценозы. Экосистемы (77)
 - [-] Ранг (13)
 - [-] Биосфера (5)

выбрать уровни онтологии | очистить

найдено 8 документов

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

1. Биоресурсный потенциал географических ландшафтов северо-запада таежной зоны России (на примере Республики Карелия)
2. Болотные экосистемы севера Европы: разнообразие, динамика, углеродный баланс, ресурсы и охрана
3. Географическая структура лесных ценофлор заповедника "Кивач"
4. Гидробиологическая характеристика водных экосистем особо охраняемых природных территорий республики Карелия
5. Нетто-фотосинтез растений, как экологический показатель биоразнообразия
6. Основы ландшафтной экологии европейских таежных лесов России
7. Фундаментальные и прикладные проблемы ботаники в начале XXI века: Материалы всеросс. конф. (22–27 сентября 2008 г.). Часть 3: Молекулярная систематика и биосистематика. Флора и систематика высших растений и флористика. Палеоботаника. Культурные и сорные р
8. Фундаментальные и прикладные проблемы ботаники в начале XXI века: Материалы всеросс. конф. (22–27 сентября 2008 г.). Часть 4: Сравнительная флористика. Урбанофлора

записать результаты в портфель | выделить все | убрать выделение

Рис. 2. Поиск по таксономии, соответствующий рубрике «БИОЛОГИЯ»

Это преимущество обусловлено, на наш взгляд, следующими основными причинами. Во-первых, массив электронных публикаций, в котором осуществляется поиск, предварительно систематизирован по предметному рубрикатору (в нашем случае по ГРНТИ). Во-вторых, индекс каждой публикации автоматически формируется с использованием таксономии терминов и, на наш взгляд, более детально характеризует ее содержание по сравнению со списком ключевых слов. В-третьих, в системе предусмотрена возможность (с использованием таксономии терминов) автоматического расширения смысла многозначных терминов запроса (например, полисемия терминов устраняется в процессе построения запроса за счет «отсечения» других предметных областей). Таким образом, разработанная процедура построения запроса в ИАС позволяет пользователю выбрать по таксономии нужные термины, а система автоматически сформирует соответствующее логическое условие отбора данных и выдаст ему нужные публикации.

В настоящее время также становится актуальной проблема разработки эффективной технологии доступа пользователей, с одной стороны, по выделенному участку карты к семантически связанным с этим участком электронным публикациям и, с другой стороны, по найденным электронным публикациям к соответствующим участкам карты, к географическим объектам которых они имеют прямое отношение. Можно отметить ряд подходов к решению этой проблемы, например, [17]. В нашем случае проблема обеспечения более «тесной» интеграции картографического сервиса ИАС с коллекциями электронных научных публикаций решается на основе разработанной технологии систематизации и поиска данных с применением онтологии, суть которой заключается в следующем.

Анализ достаточно представительного массива электронных научных публикаций по ресурсоведческим направлениям исследований (биология, лесное хозяйство, геология, водные ресурсы) показал, что в подавляющем большинстве текстов присутствуют названия географических объектов (названия населенных пунктов, болот, озер, рек, месторождений и т.п.), с которыми связаны результаты проведенных исследований. Это обстоятельство дало нам основания для применения разработанной технологии для индексации электронных публикаций по ТЕЗАУРУСУ (таксономии) ГЕОГРАФИЧЕСКИХ НАЗВАНИЙ региона и, дополнительно к «содержательному индексу» каждой публикации, приписать к нему т.н. «географический индекс». Тезаурус географических названий региона должен иметь иерархическую (древовидную) структуру (т.е. соответствовать административно-территориальному делению региона с включением названий природно-культурных и др. объектов, расположенных на данной территории, а также их синонимов, с помощью которых, на наш взгляд, можно решить вопрос, связанный с историческим переименованием географических объектов). Следует отметить, что «географический индекс» электронной публика-

ции формируется по той же схеме, что и «содержательный индекс» публикации.

Такой подход к решению задачи обеспечения более «тесной» интеграции картографического сервиса ИАС с коллекциями электронных научных публикаций позволяет реализовать следующую схему взаимодействия сервисов доступа к ГИС-данным и доступа к электронным научным публикациям.

- Пользователь с помощью сервиса доступа к ГИС-данным выделяет на тематической карте интересующую его область (например, в виде прямоугольника). Система осуществляет обратное геокодирование (т.е. реализуется процедура запроса всех названий объектов по указанным координатам с использованием ТЕЗАУРУСА ГЕОГРАФИЧЕСКИХ НАИМЕНОВАНИЙ). Далее пользователь ИАС может осуществить поиск электронных публикаций, семантически связанных с выделенными названиями географических объектов, двумя следующими способами:

1. Запустить процедуру поиска электронных публикаций, в географических индексах которых присутствуют названия указанных географических объектов.

2. Обратиться к сервису доступа к электронным публикациям и с его помощью получить нужные публикации как по интересующей его тематике, так и с учетом их связи с выделенными географическими объектами (в этом случае система при формировании логического условия отбора данных автоматически добавит в условие отбора дизъюнкцию названий соответствующих географических объектов).

- Пользователь с помощью сервиса доступа к электронным публикациям находит нужные ему по тематике публикации и при необходимости может запросить те участки карты, с которыми публикация семантически связана (в этом случае система осуществляет прямое геокодирование с использованием ТЕЗАУРУСА ГЕОГРАФИЧЕСКИХ НАИМЕНОВАНИЙ региона и по географическому индексу находит соответствующие участки тематической карты).

Таким образом, реализация предлагаемого подхода позволит пользователям ИАС находить научную информацию, которая относится к конкретному географическому региону.

Заключение

Применение методов онтологического моделирования при разработке технологий систематизации и поиска электронной научной информации в ИАС представляется, на наш взгляд, перспективным подходом при создании компьютерных систем информационной поддержки в том числе и междисциплинарных научных исследований. Основные преимущества данного подхода связаны с улучшением качества информационного поиска (полнота, точность, пертинентность) за счет: указания пользователем в запросе «правильных» терминов (выбор терминов осуществляется по таксономии); возможности авто-

матического формирования соответствующих логических условий отбора данных; устранения многозначности используемых в запросе терминов и решения вопросов синонимии. Также следует отметить возможность решения в среде ИАС задачи обеспечения интеграции картографического сервиса системы с коллекциями электронных научных публикаций.

Основные трудности связаны с разработкой онтологии (таксономии терминов) предметной области (в нашем случае таксономии терминов по тем областям науки, исследования по которым проводятся в институтах КарНЦ РАН), а также формированием логических условий предметизации, описывающих содержание электронных научных публикаций. Для решения этих сложных задач требуется привлечение высококвалифицированных

специалистов-предметников. Также следует отметить отсутствие «подходящего» для наших целей ТЕЗАУРУСА ГЕОГРАФИЧЕСКИХ НАЗВАНИЙ региона.

Одним из направлений нашей дальнейшей работы является исследование возможности использования в среде ИАС ранее созданных научных англоязычных онтологий и оценка эффективности технологий систематизации и поиска информации на соответствующем наборе англоязычных научных публикаций.

Исследовательский прототип разрабатываемой системы, реализующий часть указанных сервисов, представлен на сайте <http://ias.krc.karelia.ru>.

Авторы выражают благодарность В.Г. Старковой и Н.Б. Луговой за реализацию предлагаемых технологий и сопровождение системы.

Литература:

1. Н.Н. Добрецов, И.И. Болдырев, Р.Д. Юсупов. Гибридные информационные системы для поддержки междисциплинарных исследований // Вычислительные технологии. Том 12, Специальный выпуск 3, 2007. - С. 29-41.
2. Титов А.Ф., Вдовицын В.Т., Лебедев В.А., Полин А.К. Информационно-аналитическая система поддержки и сопровождения исследований природных ресурсов региона // Труды XII Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». - RCDL'2010, Казань. 13-16 октября 2010 г. - С. 529-534.
3. Jaudete Daltio, Claudia Bauzer Medeiros Aonde: An ontology Web service for interoperability across biodiversity applications // Information Systems. - 33 (2008). - P. 724-753.
4. Hans-Michael Muller, Eimear E. Kenny, Paul W. Stember Textpresso: An ontology-based information retrieval and extraction system for biological literature / PLoS Biology 2 (11) (2004).
5. А.Я. Гладун, Ю.В. Рогушина. Применение тезауруса предметной области для повышения релевантности поиска в интернете // Искусственный интеллект. - 2005. - № 4. - С. 742-752. [Электронный ресурс]. - URL: www.iai.dn.ua/public/JournalAI_2005_4/Razdel8/02_Gladun_Rogushina.pdf.
6. 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. - С. 70-79.
7. Россеева О.И., Загоруйко Ю.А. Организация эффективного поиска на основе онтологий. [Электронный ресурс]. - URL: http://www.dialog-21.ru/Archive/2001/volume2/2_49.htm.
8. Д.Е. Пальчунов. Решение задачи поиска информации на основе онтологий // Бизнес-информатика. - 2008. - № 1. - С. 3-13.
9. Raquel Trillo, Laura Po, Sergio Ilarri, Sonia Bergamaschi, Eduardo Mena Using semantic techniques to access web data // Information Systems. - 36 (2011). - P. 117-133.
10. David Vallet, Miriam Fernández, and Pablo Castells. An Ontology-Based Information Retrieval Model / Universidad Autynoma de Madrid Campus de Cantoblanco / Tomás y Valiente 11, 28049 Madrid.

11. Kurt Sandkuhl, Alexander Smirnov, Vladimir Mazalov, Vladimir Vdovitsyn, Vladimir Tarasov, Andrew Krizhanovsky, Feiyu Lin, Evgeny Ivashko. Context-Based Retrieval in Digital Libraries: Approach and Technological Framework // Proceedings of the 11th All-Russian Research Conference «Digital Libraries: Advanced Methods and Technologies, Digital Collections». - RCDL'2009, Petrozavodsk, Russia, 2009. - P. 151-157.
12. 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.
13. В.Т. Вдовицын, В.А. Лебедев. Онтологическое моделирование контента электронной библиотеки КарНЦ РАН // Труды КарНЦ РАН. Серия «Математическое моделирование и информационные технологии». - 2010. - № 3. Вып. № 1. - С. 11-19.
14. В.Т. Вдовицын, В.А. Лебедев. Технологии систематизации и поиска электронной научной информации с применением онтологий // Информационные ресурсы России. - 2010. - № 5. - С. 6 -10.
15. В.Т. Вдовицын, А.К. Полин. Разработка и развитие картографического сервиса информационно-аналитической системы «Природные ресурсы Карелии» // Интернет и современное общество. Сборник научных статей. Материалы XIV Всероссийской объединенной конференции «Интернет и современное общество». Санкт-Петербург, 12-14 октября 2011 г. - С. 36-39.
16. В.Т. Вдовицын, В.А. Лебедев. Оценка эффективности технологий систематизации и поиска электронной научной информации в ИАС «Природные ресурсы Карелии» // Труды 13-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». - RCDL'2011, Воронеж, Россия, 19-22 октября 2011 г. - С. 309-316.
17. Д.М. Скачков, О.Л. Жижимов. Об использовании ретроспективного геокодирования для географического поиска в электронных библиотеках // Труды 13-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». - RCDL'2011, Воронеж, Россия, 19-22 октября 2011 г. - С. 51-58.
18. Najork, M. High-Performance Web Crawling / M. Najork, A. Heydon // Kluwer Academic Publishers. - MA, USA. - 2002. - P. 25-45. [Электронный ресурс]. - URL: <http://sw.deri.org/2008/01/webcontentsurvey/paper/paper.pdf>.