Оценка эффективности технологий систематизации и поиска электронной научной информации в ИАС «Природные ресурсы Карелии»♣

© В.Т. Вдовицын, В.А. Лебедев

Институт прикладных математических исследований
Карельского научного центра РАН
vdov@krc.karelia.ru

Аннотация

В статье представлен подход к созданию и развитию информационно-аналитической системы (ИАС) поддержки и сопровождения научных исследований природных ресурсов региона. Основное внимание уделено вопросам оценки эффективности разработанных технологий систематизации и поиска электронной научной информации в ИАС с применением онтологий.

1. Введение

В настоящее время разработки информационных систем для поддержки исследований в различных областях науки и техники активно проводятся как у нас в стране, так и за рубежом [1,2,13,17]. При построении такого рода систем особую актуальность приобретают проблемы разработки и применения эффективных методов систематизации и поиска разнородной (научные публикации, тематические коллекции документов, базы данных, ГИС-системы и т.п.) электронной информации. 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) [5,7,8,12,14–16,18]. Такие системы информационного поиска учитывают смысловое содержание терминов запроса, используют онтологии, как для индексации информационных ресурсов, так и для организации семантического поиска.

В данной работе предлагаются технологии систематизации и поиска электронной научной информации, разработанные и реализованные при построении и развитии ИАС «Природные ресурсы Карелии», а также приводятся результаты проведенных экспериментов для оценки эффективности этих технологий. Наиболее близкими по теме наших исследований и разработок являются подходы, представленные, например, в работах [2,8,15].

2. 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 Веб-краулера [4], который в процессе своей работы формирует в ИАС временное хранилище электронных научных публикаций, карт, космических снимков и соответствующих

Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2011, Воронеж, Россия, 2011.

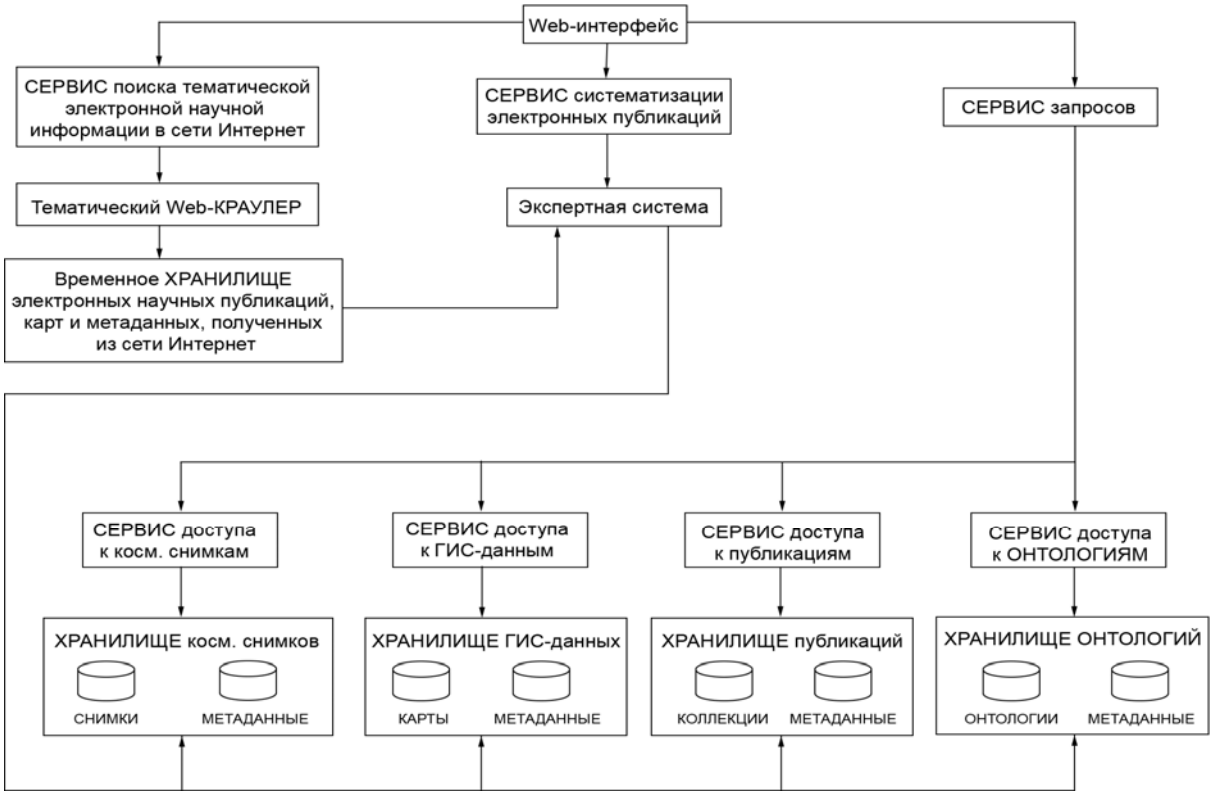

Рис.1 Архитектура ИАС «Природные ресурсы Карелии» – основные компоненты

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 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 запроса к ИАС, которая включает: выбор ключевых слов с использованием онтологии; формирование логического условия отбора данных с использованием списка ключевых слов и логических операций – AND, OR, NOT; выполнение процедуры поиска в хранилищах системы с использованием логического условия отбора данных и базы индексов публикаций; автоматическое ранжирование результатов поиска по степени релевантности; 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.

3. Технологии систематизации и поиска в ИАС «Природные ресурсы Карелии»

В рамках создания 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 [6,9–11].

Онтологию можно определить как набор формализованных явных описаний терминов предметной области и отношений между ними (Gruber, T.R.). В нашем случае процедура формирования предметной онтологии по выбранным направлениям естественных наук заключается в следующем. По ресурсоведческим направлениям исследований, развиваемым, в частности, в КарНЦ РАН, устанавливается список научных дисциплин и предметов их изучения. По каждому предмету составляются списки морфологических признаков и анатомических частей, списки свойств, списки классификаций и классов по свойствам, списки взаимодействий и воздействий между парами предметов (систем) и их классов. Далее, устанавливаются парадигматические отношения между терминами и, в соответствии с ними, строится иерархическая структура связей терминов. Формирование предметной онтологии проводится с привлечением ведущих ученых Центра и с учетом ранее созданных в других российских и зарубежных организациях подобных предметных онтологий. Разработанная онтология представляется в виде базы данных. Для поддержки процессов создания и сопровождения онтологий разработаны соответствующие программные сервисы.

Систематизация 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 используется ряд номенклатур из таксономии терминов.

Для разработки логических условий нами проанализирован достаточно представительный массив научных публикаций сотрудников КарНЦ РАН. В результате проведенного анализа и консультаций специалистов были определены следующие типы публикаций по характеру работ безотносительно 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>).**

Термины в угловых скобках (нетерминалы) символически представляют номенклатуры терминов, являющихся частью таксономии терминов. Предполагается, что для каждой рубрики ГРНТИ это будут свои номенклатуры (хотя одна и та же номенклатура может входить в условия разных рубрик); AND, OR и NOT – логические операции конъюнкции, дизъюнкции и отрицания. Основное требование к логическим условиям заключается в том, что они должны содержать все номенклатуры терминов, определяющие содержание соответствующей рубрики, чтобы не «потерять» релевантные публикации. Следует отметить, что на первом этапе процесса систематизации (этапе предметизации) могут быть предметизированы и нерелевантные публикации. На втором этапе (этапе индексации) предметизация уточняется.

Следует также 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 составляться как комбинация из указанных схем. 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. По аналогичной схеме выполняется предметизация по всем имеющимся в информационной системе коллекциям публикаций. При этом метаданные публикаций помечаются для того, чтобы не предметизировать их повторно (например, после очередного пополнения коллекций системы новыми публикациями). Публикации, для которых попытка предметизации не дала результата, подвергаются повторной предметизации при каждом запуске ЭС в расчете на возможное пополнение/корректировку набора логических условий предметизации (т.е. правил-продукций ЭС).

В процессе индексации в тексте каждой публикации ищутся термины соответствующего рубрике фрагмента таксономии, и определяется их место в таксономии. Если при этом находится пара

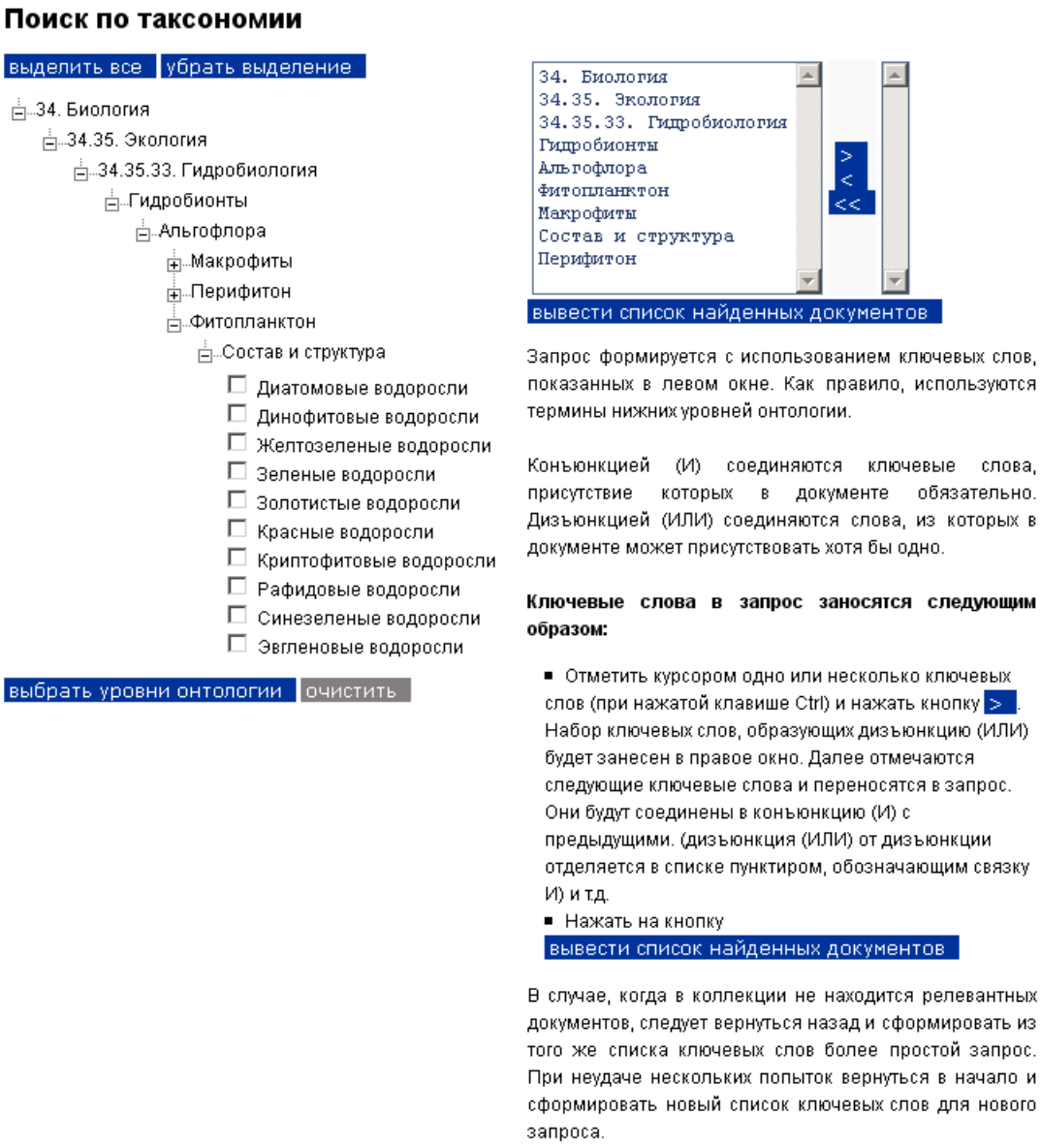

Рис. 2. Поиск по таксономии, соответствующий рубрике «БИОЛОГИЯ»

терминов, лежащих на одной ветви таксономии (за исключением пар, содержащих название рубрики), то эта ветвь помещается в индекс. Так же происходит со всеми найденными в тексте публикации терминами. В результате индекс 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. Простейший вид запроса состоит в требовании найти 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 сохранения текстов публикаций в «личном» кабинете пользователя.

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 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 этих терминов формулируется запрос в виде логического выражения, определяющего конъюнктивные и дизъюнктивные связи терминов. 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. Вид выражения выводится на экран для того, чтобы пользователь мог его оценить и скорректировать в случае необходимости.

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.

4. Оценка эффективности работы алгоритмов систематизации и поиска

Эффективность системы поиска информации характеризуется следующими основными показателями: **полнота, точность, пертинентность**, а также затратами 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 этого времени целесообразно систематизировать массив публикаций так, чтобы поиск выполнялся только в определенной части массива. **Точность** поиска означает, что в отклике на запрос присутствуют именно те публикации, которые содержат наборы терминов запроса. Пертинентность отклика означает, что отобранные релевантные запросу публикации соответствуют информационным потребностям пользователя, его специальности, области интересов и, в идеальном случае, не содержат публикации из других предметных областей. Одним из средств «борьбы» за точность и пертинентность поиска является систематизация публикаций. Кроме того, повышению полноты и точности поиска способствует технология построения запросов, основанная на соответствующей систематизации предметных областей. Существенно сокращает время поиска индексация текстов публикаций. В этом случае вместо полнотекстового поиска по всему массиву публикаций выполняется поиск в базе индексов, что существенно быстрее.

Для измерения эффективности методов информационного поиска используется тестовый набор данных, на котором строится оценка качества [3]. Данный набор включает:

1. тестовую коллекцию документов;
2. тестовое множество информационных потребностей пользователя, выражаемых в виде запросов;
3. набор бинарных оценок для каждого найденного документа, характеризующих релевантность или нерелевантность данных документов к запросам.

Для проведения экспериментов по оценке качества информационных технологий систематизации и поиска информации в ИАС нами выбрана тестовая коллекция электронных научных публикаций в области биологических наук в количестве 1000 документов.

Для исследования качества предлагаемых методов систематизации и поиска информации в ИАС использовались традиционные метрики**:** полнота – **r = a/(a+c)**; точность – **p = a/(a+b) (**где: **a** – найденные релевантные документы, **b** – найденные нерелевантные документы, **c** – ненайденные релевантные документы). Также мы используем показатель пертинентности, который определяется отношением количества релевантных документов, отнесенных к рубрике, соответствующей специальности или области интересов пользователя (то есть к рубрикам ГРНТИ, которые пользователь выбрал сам) к общему количеству документов в отклике на запрос [15]. Выбор рубрики предоставляет наша технология, а стандартные интернет-поисковики выполняют поиск по всему массиву документов. В результате в отклике появляются документы из областей, не интересующих пользователя, и пертинентность отклика падает. Например, на запрос «альгофлора», заданный гидробиологом, Яндекс выдает статьи по альгофлоре почв и болот, что не соответствует потребностям гидробиолога. В то же время Яндекс не может раскрыть объем понятия альгофлора, и ищет в текстах только этот термин. Наша технология подразумевает выбор области интересов (то есть предметных рубрик) до начала построения запроса. При этом запрос автоматически расширяется терминами, раскрывающими «объем» термина исходного запроса. Тем самым обеспечивается полнота отклика на запрос и достигается значение пертинентности близкое к единице. Обозначим пертинентность через **P**. Тогда, **P = a1/(a+b)**, где: **a1** – количество пертинентных документов, **(a+b)** – общее количество документов в отклике на запрос. Для оценки качества ранжирования результатов запроса по степени

Таблица 1

	Запрос	Поиск по Яндексу			Поиск по онтологии		
		r	**p**	**P**	**r**	**p**	**P**
1	Недревесные лесные ресурсы	0,40	0,47	0,30	0,74	0,94	0,78
2	Альгофлора	0,41	0,68	0,61	0,89	0,77	0,77
3	Лекарственные растения	0,60	1,00	1,00	1,00	1,00	1,00
4	Паразиты рыб	0,90	0,91	0,91	0,91	0,98	0,98
5	Действие физических факторов на растения	0,68	0,42	0,35	0,68	0,93	0,86
6	Лесоводство. Методы ухода	0,71	0,63	0,53	0,82	1,00	1,00
7	Наземные позвоночные. Болезни, паразиты	0,46	0,41	0,4	0,88	0,95	0,95

релевантности информационным потребностям пользователя можно использовать следующую метрику: **Precision (n) = k/n** (**k** – количество релевантных документов среди первых **n** – документов отклика.

Для оценки эффективности поиска с использованием онтологии выполнены эксперименты по поиску по ключевым словам (с использованием поисковика Яндекс) и по разработанной нами технологии.

Для проведения экспериментов были выбраны следующие запросы:

- Недревесные лесные ресурсы.
- Альгофлора.
- Лекарственные растения.
- Паразиты рыб.
- Действие физических факторов на растения.
- Лесоводство. Методы ухода.
- Наземные позвоночные. Болезни, паразиты

4.1 Оценка эффективности алгоритмов предметизации и индексации

Эффективность алгоритмов предметизации и индексации определяется полнотой и точностью сформированных логических условий предметизации (правил-продукций ЭС) и таксономии терминов, определяющих содержание соответствующей предметной рубрики. На данный момент проведения исследований сформированы таксономии терминов и логические условия предметизации для 32 предметных рубрик ГРНТИ. После выполнения предметизации и индексации тестового массива документов были проведены оценки релевантности их результатов. Анализ массива предметизированных документов показал, что все они релевантны соответствующим предметным рубрикам. После этого был проанализирован «остаток» непредметизированных документов тестового массива публикаций. По разным рубрикам в «остатке» было обнаружено от 0 до 26% документов, релевантных соответствующим рубрикам. Это были документы, не содержавшие терминов рубрик ни в названиях, ни в списках ключевых слов. В основном это были сборники статей и отдельные статьи обзорного характера. Релевантные статьи, содержащиеся в этих сборниках, в большинстве случаев, были уже ранее предметизированы как отдельные статьи. Обобщающие статьи, хотя и содержат релевантные термины, но содержание их текстов, как правило, не отличается новизной. Тем не менее, была выполнена коррекция логических условий предметизации, в результате которой статьи из «остатка» оказались предметизированы по соответствующим рубрикам ГРНТИ. Однако пока еще нет полной гарантии полноты предметизации, поэтому в технологию предметизации и индексации включен дополнительный этап – индексация «остатков» по всей имеющейся таксономии терминов, что, на наш взгляд, гарантирует полную предметизацию всех электронных публикаций при условии полноты самой таксономии.

4.2 Оценка эффективности технологии поиска в ИАС

Технология поиска основана на булевской модели оценки релевантности. Как было указано выше, для оценки эффективности технологии поиска с использованием таксономии терминов было сформировано 7 запросов разной степени сложности с целью оценки качества поиска, как по сравнению с работой поисковика Яндекс, так и по предложенной нами технологии поиска. Результаты проведенных экспериментов сведены в таблице 1.

Из таблицы 1 видно, что в случае запросов, содержащих многозначные термины, поисковик Яндекса выдает довольно скромные результаты в плане полноты отклика на запрос за исключением отклика на запрос **паразиты рыб**. Последнее объясняется тем, что словосочетание **паразиты рыб** встречается в заголовках и текстах почти всех

имеющихся релевантных запросу публикаций. Словосочетание **лекарственные растения** встречается примерно в 90% публикаций, а словосочетание **недревесные лесные ресурсы** появляется в 40% публикаций. При поиске с использованием онтологии эти запросы автоматически расширяются включением в их состав терминов, раскрывающих содержание терминов запроса. Например, **недревесные лесные ресурсы** включает термины: **пищевые** (и список – ягод, грибов и орехов), **лекарственные** (и список лекарственных), **рекреационные** (и список – туризм, охота, рыбалка). Предварительно определено, что запрос **альгофлора** задан гидробиологом, поэтому публикации по альгофлоре почв и болот для него **не пертинентны** (отсюда следует довольно низкая оценка пертинентности поиска Яндексом). Низкая оценка полноты объясняется тем, что Яндекс не имеет информации об объеме и содержании понятия **альгофлора** (в таксономии термин **альгофлора** включает термины: фитопланктон, перифитон, макрофит, а также списки видов, входящих в их состав). В других случаях оценки пертинентности довольно высокие. Это объясняется тем, что в запросах использованы однозначные термины. Следует отметить, что во многих случаях запросы могут содержать многозначные термины и тогда оценки пертинентности отклика на запрос в поисковике Яндекса могут резко упасть.

Из рассмотренных примеров можно сделать предварительный вывод о том, что эффективность поиска с использованием онтологии, как и ожидалось нами, 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.

5. Заключение

Таким образом, предварительные результаты проведенных экспериментов для оценки эффективности разработанных и реализованных технологий систематизации и поиска электронных публикаций в ИАС показали перспективность предлагаемого подхода. В настоящее время эти технологии реализованы в ИАС не в полном объеме (предметизация электронных публикаций проводилась только по их названиям и без учета соответствующих списков ключевых слов, требуется доработка (расширение, уточнение) предметных онтологий и логических условий предметизации, также пока не реализован и механизм ранжирования публикаций). Тем не менее, они превзошли по качеству поиска Яндекс. 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). Процедура построения запроса в ИАС позволяет пользователю выбрать по таксономии нужные термины для формирования логического условия отбора данных. При этом даже если пользователь указывает в запросе только один термин, запрос перед исполнением автоматически пополняется терминами названий предыдущих и последующих уровней таксономии (и их синонимами), которые он прошел до выбора интересующего его термина.

Исследовательский прототип разрабатываемой системы, реализующий большую 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] Jaudete Daltio, Claudia Bauzer Medeiros Aonde: An ontology Web service for interoperability across biodiversity applications //Information Systems 33 (2008) P. 724–753.

[2] Hans-Michael Muller, Eimear E. Kenny, Paul W. Sternber Textpresso: An ontology-based information retrieval and extraction system for biological literature / PLoS Biology 2 (11) (2004).

[3] Manning, C. An Introduction to Information Retrieval / Christopher D. Manning, Prabhakar Raghavan, Hinrich Schьtze. – Cambridge, England: Cambridge University Press. – April 2009. – P. 544 (84–133, 151–217, 443–481).

[4] Najork, M. High-Performance Web Crawling / M. Najork, A. Heydon // Kluwer Academic Publishers. – MA, USA. – 2002. pp. 25–45. http://sw.deri.org/2008/01/webcontentsurvey/paper/paper.pdf

[5] Roberto Navigli Word Sense Disambiguation: A Survey // ACM Computing Surveys, Vol. 41, No. 2, Article 10, Publication date: February 2009, 69 pages DOI = 0.1145/1459352.1459355 http://doi.acm.org/10.1145/1459352.1459355

[6] 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.
[7] Raquel Trillo, Laura Po, Sergio Ilarri, Sonia Bergamaschi, Eduardo Mena Using semantic techniques to access web data //Information Systems. 36 (2011). P. 117–133.
[8] David Vallet, Miriam Fernбndez, and Pablo Castells An Ontology-Based Information Retrieval Model /Universidad Autynoma de Madrid Campus de Cantoblanco / Tombs y Valiente 11, 28049 Madrid
[9] 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.
[10] В.Т. Вдовицын, В.А. Лебедев «Онтологическое моделирование контента электронной библиотеки КарНЦ РАН» //Труды КарНЦ РАН, № 3. 2010. Серия «Математическое моделирование и информационные технологии». Вып. № 1. С. 11–19.
[11] В. Вдовицын, В. Лебедев Технологии систематизации и поиска электронной научной информации с применением онтологий //Информационные ресурсы России. – 2010. – № 5. – С. 6 –10.
[12] А.Я. Гладун, Ю.В. Рогушина Применение тезауруса предметной области для повышения релевантности поиска в Интернете //«Искусственный интеллект» 4'2005. С.742–752 – www.iai.dn.ua/public/JournalAI_2005_4/Razdel8/02_Gladun,_Rogushina.pdf
[13] 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.
[14] 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.
[15] Д.Е. Пальчунов Решение задачи поиска информации на основе онтологий //Бизнес информатика № 1–2008 г. С. 3–13.
[16] Россеева О.И., Загорулько Ю.А. Организация эффективного поиска на основе онтологий – http://www.dialog-21.ru/Archive/2001/volume2/2_49.htm
[17] 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.
[18] Труды Симпозиума «Онтологическое моделирование». //Под ред. Л.А. Калиниченко. – М.: ИПИ РАН, 2008. – 303 с.

Evaluation of the Technology Effectiveness of the Systematization and Search of Digital Scientific Information in the IAS «Natural Resources of Karelia»

© Vladimir Vdovitsyn, Viktor Lebedev

The article presents an approach to the creation and development of information-analytical system (IAS) of support and maintenance of scientific research of natural resources in the region. It emphasizes the assessment of the effectiveness of the developed technologies of the systematization and search of e-science information in the IAS using ontologies.

♣ The paper is based on research carried out as a part of the project CoReLib supported by the Swedish Institute by grant # 00760-2010