

# Онтологии для тематического поиска данных в коллекциях электронной библиотеки\*

© В.Т. Вдовицын, В.А. Лебедев

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

## Аннотация

В работе рассматриваются вопросы практического построения и применения онтологий для организации тематического поиска данных в коллекциях электронной библиотеки. Представлен исследовательский прототип системы сервисов, обеспечивающих построение и редактирование онтологий, контроль целостности коллекций, а также формирование запросов и поиск релевантных документов.

## 1 Введение

Электронная библиотека Карельского научного центра РАН (ЭБ КарНЦ РАН) – <http://dl.krc.karelia.ru> предназначена для формирования, хранения, сопровождения и поиска электронных коллекций научных информационных ресурсов, отображающих результаты исследований природы и культурно-исторического наследия Карелии. В настоящее время ЭБ содержит коллекции по сосудистым растениям, афиллофороидным грибам, биотопам, млекопитающим, минеральным ресурсам и др. Кроме того создана и пополняется коллекция научных публикаций сотрудников, изданных редакционно-издательским отделом КарНЦ РАН. Предполагается дальнейшее развитие ЭБ как в плане расширения состава и содержания коллекций, так и совершенствования программных сервисов поддержки процессов формирования, хранения, сопровождения и поиска данных.

В предшествующих публикациях [4, 6, 7] изложены технологические основы построения ЭБ с применением XML. В данной работе основной акцент сделан на развитии сервиса поиска данных с применением онтологий.

Первоначально доступ к данным коллекций ЭБ осуществлялся выбором из списков названий документов в коллекции или набором на клавиатуре названий некоторых рубрик документов. Впоследствии выяснилась актуальность тематического подбора документов из разных коллекций по некоторому поисковому запросу, формирование которого не

должно представлять затруднений для пользователя. Таким образом, мы подошли к проблеме применения онтологий для поиска. Различным аспектам применения онтологий посвящен ряд наших работ [5, 11, 12], а также работы других авторов [2, 8–10, 13, 15]. Например, в обзорной работе [2] представлена, в частности, классификация моделей и языков для конструирования онтологий, а в обзоре [13] рассматривается использование онтологий для поддержки задач интеграции в семантически гетерогенных информационных системах и приводится сравнительный анализ инструментов.

Данная работа посвящена анализу состава, структуры и применения онтологий в ЭБ КарНЦ РАН.

## 2 Состав онтологий

Как известно, понятие «онтология» ввел в научный оборот информатики Т. Gruber [1], определив его как «точная спецификация концептуализации» предметной области. В дальнейшем в связи со сложностью реальных предметных областей и соответственно затруднительностью построения универсальных онтологий М. Uschold и М. Grundiger [3] предложили добавить в это определение уточнение – «с учетом задач информационной системы».

Назначением ЭБ КарНЦ РАН является представление в сети Интернет материалов по изученности свойств, состояния ресурсов региона в виде набора коллекций, содержащих описания классов однородных объектов, например таких, как минеральные ресурсы, водные объекты, виды животных и растений. При этом каждый такой класс определяется соответствующим понятием.

Понятие обозначается термином и является представлением класса реальных или абстрактных объектов и обладает смыслом и значением. В компьютерной информатике используется только значение, которым является совокупность (агрегат) свойств, атрибутов, признаков, характеристик класса объектов. Свойство (атрибут) – абстрактный объект, значение которого выбирается из некоторого именованного множества (домена). При этом домены могут быть различного типа: числовые, символьные, графические (рисунки, графики), составные и др.

Свойства и характеристики реальных объектов могут изучаться различными научными дисциплинами и играть различные роли в их описании или использовании. В соответствии с этим свойства мо-

---

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008.

гут различными способами группироваться (агрегироваться), что порождает иерархическую тематическую структуру над свойствами. Тем самым вводятся дополнительные понятия. Помимо этого, как правило, однородность объектов в составе класса относительна, что часто вызывает необходимость выделения подклассов, то есть введения дополнительных понятий.

Таким образом, описание предметной области (области знаний) содержит некоторый набор понятий, которые вступают между собой в различные отношения: подчинения, включения, пересечения, ассоциации и др.

Отношения между понятиями обычно сводятся к небольшому числу классов отношений между терминами: классификации, агрегации, ассоциации. Классификации различаются по основаниям классифицирования (по их названиям), агрегации – по названиям целого. Среди ассоциаций важную роль играет отношение равнозначности, выражающее полную или частичную синонимичность терминов. Обычно применяются две схемы выражения бинарных отношений между терминами:  $ARB$  или  $R(A,B)$ , где:  $R$  – имя отношения;  $A, B$  – термины.

Отношения между объектами предметной области обычно конкретны и представляются в виде некоторой характеристики класса объектов, в которой присутствуют выражения, содержащие термины, обозначающие другой класс объектов (реальных или абстрактных) или его характеристики. Такие характеристики задаются функциями, в частности, ролевыми функциями. Последние выражают роли некоторого класса объектов по отношению к объектам другого класса. Например, в экосистемах одни животные по отношению к другим могут быть врагами, конкурентами и т.п. Отметим, что совокупность конкретных отношений определяется назначением создаваемой информационной системы для решения заданного класса задач.

Для пояснения сказанного выше рассмотрим пример. Пусть необходимо создать систему «Определитель видов растений». Систематическая классификация растений основана на классификациях их морфологических признаков: подземных органов (корней, клубней и др.), вегетативных органов (стеблей, листьев), генеративных органов (цветков, соцветий, плодов и др.), то есть, на основе выделения таксонов по форме и размерам соответствующих органов. Каждый таксон именуется многословным термином, составляющими которого являются термины классификаций признаков данного органа по различным основаниям.

Существует несколько алгоритмов идентификации видов по выделенным признакам. В основном они схожи: выделяются таксоны ряда органов для определения класса, затем рассматриваются дополнительные таксоны для определения семейства внутри класса и, наконец, наиболее характерные таксоны используются для идентификации вида. Таким образом, процесс определения вида выполняется в два этапа – производится сравнение форм идентифици-

руемого растения с имеющимися в справочнике описаниями растений, затем формируется ответ – название вида растения (см. например [14]). «Не подготовленный» пользователь, как правило, не владеет терминологией, идентифицирующей, например, формы морфологических органов растений, поэтому целесообразно дополнить словарь терминов базой данных, в которой каждому термину ставится в соответствие изображение органов растений. Термины и изображения группируются в соответствии с иерархией, а также по стадиям вегетации растений. База данных должна содержать признаки каждого вида растений, сгруппированные по частям растений и стадиям вегетации. Таким образом, для решения задачи определения вида при помощи некоторой информационной системы на наш взгляд необходимо создать соответствующую прикладную онтологию, включающую систему формирования и исполнения запросов и базу данных, содержащую совокупность признаков каждого вида растения.

В данной работе предлагается система тематического поиска данных в коллекциях ЭБ, которая включает следующие основные компоненты:

- словарь связей терминов (таксономию) и их определений,
- сервис доступа к словарю с отбором релевантных терминов для включения в запрос,
- сервис формирования запроса в виде логического выражения (с визуальным контролем его корректности),
- сервис формирования по этому выражению оператора "Select",
- сервис поиска релевантных данных и выдачи их пользователю.

Таким образом, в состав онтологии в нашем случае включается словарь, отображающий иерархию терминов с альбомом изображений, соответствующих их названиям, и перечисленные выше программные сервисы, определяющие правила и ограничения использования этих терминов для построения пользователем тематических запросов к коллекциям ЭБ.

Предполагается, что система должна работать следующим образом. Пользователь сначала выбирает группу органов, например, подземные, вегетативные, генеративные органы. Далее для уточнения запроса ему выдается список, например, вегетативных органов: стебель, листья. А при выборе, например, листьев на экран выдается группа изображений листьев, в которой он должен сделать выбор и т.д. до формирования списка терминов, описывающих морфологию растения. На основе этого списка формируется запрос, который и подается системе исполнения запросов для поиска релевантных данных в коллекциях ЭБ и выдачи ответа.

### 3 Структура словаря

Коллекции ЭБ предназначены для описания состава, состояния и взаимодействий конкретных объектов (например, рек и озер) или классов (например, видов растений). Определение понятий, отношений и

механизмов управления, необходимых для описания процессов решения задач в избранной предметной области, выполняется специалистами предметниками в процессе разработки структур документов и баз данных. При этом выделение рубрик, атрибутов и их значений опирается на традиционную систему понятий данной предметной области. В итоге рубрикация документов и выражение содержания рубрик определяют структуру и состав словаря для данной коллекции. Как было указано выше, все понятия концептуальной схемы попарно связаны отношениями. Следовательно, словарь онтологии представляет собой совокупность триад вида  $R(A,B)$  (или  $ARB$ ), связи между которыми устанавливаются по тождеству написания названий отношений, первых или вторых терминов. В результате получается ациклический граф, в котором вершины помечены терминами, а дуги – названиями отношений.

Иерархию в этом графе определяют классы отношений классификации и агрегации. На нижних уровнях находятся термины, обозначающие характеристики и их значения. Выше располагаются агрегаты и классы в соответствии с их подчинением. Класс отношений равнозначности (синонимии) следует выделить из структуры основного графа, так как они могут проявляться на любом уровне иерархии и из-за нерегулярности излишне усложняют структуру словаря.

Среди терминов часто присутствуют многословные, например, «сухие еловые леса на сравнительно богатых почвах», «листья очередные простые яйцевидно-эллиптические тупые сидячие 0,4–2,2 см длиной, 0,2–1,1 см шириной». Наличие многословных терминов свидетельствует о многократной классификации данного класса объектов по различным основаниям. Включение многословных терминов непосредственно в состав словаря значительно усложнило бы процесс определения релевантности документов. Поэтому, на наш взгляд целесообразно включать в состав словаря каждую независимую классификацию отдельно.

Расположение независимых классификаций на одном уровне иерархии приводит к разрастанию соответствующего фрагмента словаря, делая его слабо обозримым, и затрудняет выбор нужных значений. Вследствие этого целесообразно на наш взгляд разносить их по уровням иерархии.

Различение отношений по названиям не всегда несет полезную информацию для пользователя. Связи между терминами часто настолько привычны, что для пользователя не требуется уточняющих сведений. В этом случае возникает возможность не помечать дуги графа, что значительно упрощает реализацию словаря.

Как известно, в общем виде граф  $G$  есть пара  $\langle N, E \rangle$ , где  $N$  – множество вершин,  $E$  – множество дуг. Если каждую дугу графа представить в виде пары смежных вершин, обозначенных терминами, и добавить к полученному множеству список висячих вершин, то иерархический граф словаря может быть представлен в виде «двухколонковой» таблицы, в

которой в первой колонке представлен список «вершин-родителей», а во второй – список «вершин-потомков». Такой словарь может быть реализован в виде реляционной или объектной базы данных.

Существует две возможности реализации словаря: в виде множества нормализованных таблиц или в виде единой таблицы. В первом случае обеспечивается доступ к словарю по уровням иерархии, начиная с корня (поиск «в глубину»). Во втором случае становится возможным обращение к любому слову на любом уровне иерархии, если обеспечить предварительный поиск слова по набору его начальных букв. Однако, если структура словаря не древовидна, то есть содержит «полуциклы» (а это неизбежно при наличии многословных терминов), то возникают затруднения при построении сложных запросов. Это будет видно ниже при рассмотрении структуры запросов.

Сервис доступа к словарю должен обеспечивать последовательный отбор терминов от корня к листьям. Получающийся при этом список терминов сохраняет их иерархическую соподчиненность, что позволяет строить запросы, учитывающие иерархию терминов. Тем самым отсекается большая часть «информационного шума» при поиске.

Словари рассмотренного вида определяют объем и содержание каждого термина, обозначающего понятие, то есть его значение в коллекции электронной библиотеки. Значения характеристик, за исключением тех терминов, семантика которых не требует специальных определений ввиду ее прозрачности для пользователей, могут определяться заданием соответствующих функций, рисунков или словесным определением (дефиницией) в зависимости от типа значений. Это позволяет использовать такие словари для решения задач поиска данных. Кроме того, такого рода словари могут использоваться и для контроля целостности коллекций ЭБ, то есть для выявления опечаток, неучтенных синонимов и др.

#### 4 Структура запроса

Задача состоит в том, чтобы построить на основе словаря логическое выражение, семантически и синтаксически соответствующее аргументу оператора "Select". При этом логическое выражение должно удовлетворять следующим требованиям:

- Должна сохраняться иерархическая соподчиненность терминов. Это обеспечивает отсеивание «информационного шума» в тех случаях, когда термины, во-первых, могут появляться в разных документах в различных рубриках и на разных уровнях (например, название одного вида или рода животных может относиться к рубрике «систематика» в одном документе и к рубрике «враги, конкуренты, паразиты» в другом), во-вторых, компоненты многословных терминов могут присутствовать в текстах в другом словесном окружении.
- Должно четко проводиться различие между списками терминов, присутствие которых обязательно или желательно. Например, все компо-

ненты многословных терминов являются обязательными, что должно соответственно оформляться синтаксически.

- Связь между характеристиками и их значениями должна указываться явно.

В соответствии с этими требованиями принята следующая структура запроса:

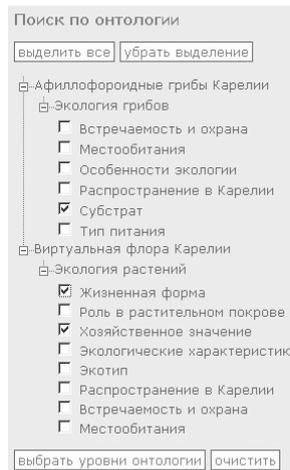
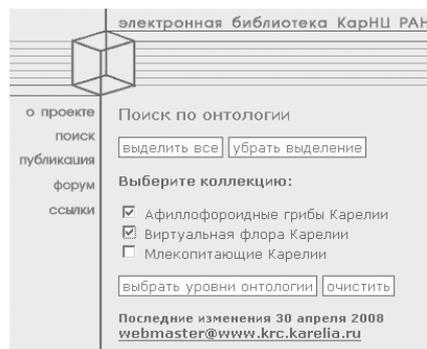
**(Вид объекта) [класс ... Л] [тема ... Л] [характеристика [= <список значений>]] <Л / V> ... <Л / V> [класс ... Л] [тема ... Л] [характеристика [= <список значений>]],**

где: <список значений> ::= <значение> | <значение> Л <список значений> | <значение> V <список значений>

Нетерминалы, заключенные в квадратные скобки, представляют термины из их множества на соответствующем уровне иерархии. Они могут присутствовать или отсутствовать в запросе в соответствии с иерархией скобок. Многоточие после нетерминалов «класс» и «тема» обозначает, что в данном списке может быть один и более терминов, образующих иерархию. Нетерминал «Вид объекта» по существу представляет множество коллекций ЭБ и определяется в момент выбора коллекции, поэтому в списке он заключен в круглые скобки. Нетерминал «класс» означает наличие некой классификации, возможно иерархической, видов объектов (например, озер, видов животных), описания которых представлены в коллекции. Нетерминал «тема» указывает наличие агрегации свойств объектов в группы, например, по отношению к научным дисциплинам, изучающим их. Объекты в коллекции представлены описаниями их свойств и значений. В зависимости от интереса пользователя в данный момент запрос может содержать те или иные термины от одного до десятка и более.

Построение запроса по указанной выше схеме представляет определенные трудности для пользователя, поэтому в системе предусмотрен сервис построения запросов.

Реализация словаря в виде совокупности взаимосвязанных таблиц, содержащих дуги графа словаря, относящиеся к одному уровню иерархии, позволяет пользователю последовательно отбирать термины в процессе «поиска в глубину». В результате этого получается список терминов, релевантный запросу пользователя. Так как в процессе поиска нужных терминов в словаре, как правило, отбираются и избыточные термины, то пользователю предоставляется возможность не включать некоторые из них в запрос. В процессе формирования «суженного» списка терминов автоматически выполняется его пополнение синонимами. Далее работает сервис формирования запроса в соответствии с указанной выше схемой и с учетом иерархической соподчиненности терминов. По этому запросу выполняется поиск, в результате которого на экран выводится список релевантных документов (возможно, пустой) и вид логического выражения запроса, что обеспечивает возможность его корректировки. Процесс формирования запроса и вид отклика показаны на рис. 1.



Запрос формируется с использованием ключевых слов, показанных в левом окне. Как правило, используются термины нижних уровней онтологии.

Конъюнкцией (И) соединяются ключевые слова, присутствие которых в документе обязательно. Дизъюнкцией (ИЛИ) соединяются слова, из которых в документе может присутствовать хотя бы одно.

Ключевые слова в запрос заносятся следующим образом:

- Отметить курсором одно или несколько ключевых слов (при нажатой клавише Ctrl) и нажать кнопку [ > ]. Набор ключевых слов, образующих дизъюнкцию (ИЛИ) будет занесен в правое окно. Далее отмечаются следующие ключевые слова и переносятся в запрос. Они будут соединены в конъюнкцию (И) с предыдущими. (дизъюнкция (ИЛИ) от дизъюнкции отделяется в списке пунктиром, обозначающим связку И) и т.д.
- Нажать на кнопку

вывести список найденных документов

В случае, когда в коллекции не находится релевантных документов, следует вернуться назад и сформировать из того же списка ключевых слов более простой запрос. При неудаче нескольких попыток вернуться в начало и сформировать новый список ключевых слов для нового запроса.

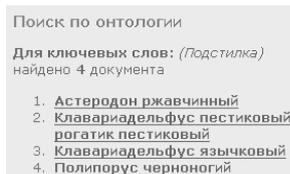


Рис. 1. Интерфейсы для формирования запроса и представления результатов поиска

## 5 Применение онтологии

Итак, целью применения онтологии в составе нашей ЭБ является обеспечение «дружественного» интерфейса с коллекциями ЭБ для организации тематического поиска объектов, чьи свойства соответствуют запросам пользователя в данный момент. При этом под термином «онтология» в нашем случае понимается словарь, отображающий иерархию терминов моделируемой предметной области с альбомом изображений, соответствующих их названиям, и перечисленные выше программные сервисы, определяющие правила и ограничения использования этих терминов для построения пользователем тематических запросов к коллекциям ЭБ.

Словарь формируется на базе ГРНТИ словарей предметных областей, описания которых представлены в коллекциях ЭБ. При этом в ряде случаев имеет место пересечение множеств терминов тех словарей, объекты коллекций которых взаимодействуют между собой (например, отражают взаимодействие организмов в экосистемах). Каждый предметный словарь является отображением концептуальной схемы описания объектов данной коллекции, принятой в соответствующей научной дисциплине.

Система сервисов формирования и исполнения запросов разрабатывается с использованием СУБД MySQL и обеспечивает не только поиск данных в коллекции, но и загрузку, редактирование словарей и некоторые другие функции, связанные с применением онтологий [5, 7, 12].

Как было показано выше, структура словаря представляет собой иерархический граф с полуциклами, в котором дуги обозначены парами смежных вершин (то есть, парами терминов). В этом случае граф словаря может быть представлен таблицей, число граф (столбцов) которой соответствует числу уровней иерархии, и которая естественным путем поддается декомпозиции.

Загрузка и декомпозиция каждого предметного словаря ЭБ, так же как и контроль его полноты и целостности (то есть, проверка связности графа, проверка отсутствия циклов и аутентичность написания терминов), выполняется автоматизированно. Эта операция реализуется посредством «поиска в глубину» экспертом-специалистом в данной предметной области, который имеет возможность самостоятельно редактировать словарь. На рис. 2 представлен интерфейс пополнения и редактирования словаря.

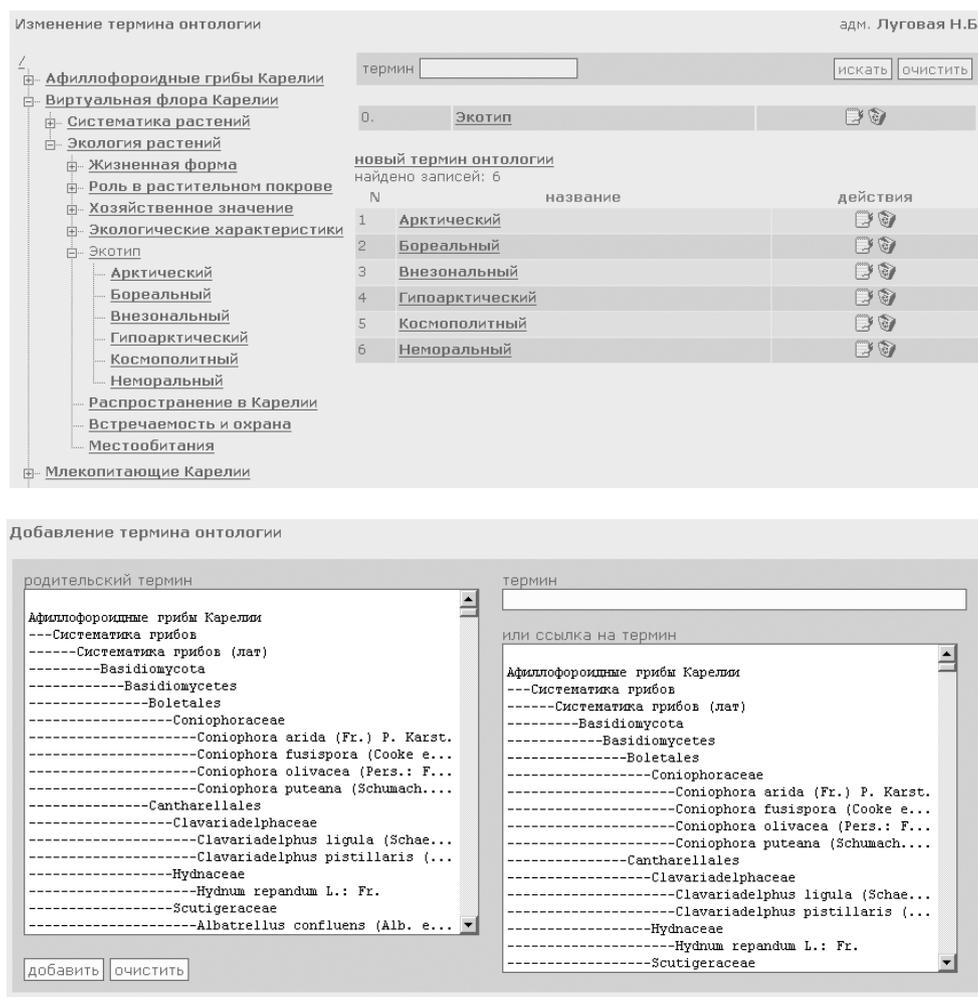


Рис. 2. Интерфейс для пополнения и редактирования словаря

Таким образом, совокупность протестированных предметных словарей объединяется в единую базу данных на основе рубрикатора ГРНТИ с целью обеспечить пользователю возможность выбора интересующей предметной области. При этом использование разработанной таким образом онтологии предоставляет пользователю ряд следующих преимуществ:

- Возможность обозреть концептуальную схему предметных областей, представленных своими описаниями в ЭБ.
- Обеспечивает отбор терминов для запроса со всеми синонимами в аутентичном написании.
- Способствует построению «правильных» запросов.
- Обеспечивает переформулирование запроса в случае получения «пустого» отклика.

Таким образом, одной из основных функций применения онтологии в ЭБ является функция построения тематических запросов на поиск данных в коллекциях ЭБ, реализуемая по схеме, указанной в предыдущем разделе.

Документы коллекций ЭБ сформированы в виде: реляционных баз данных, слабо структурированных текстов в формате XML-DTD, а также в форматах HTML и RTF. В соответствии с этим организован и поиск данных, релевантных запросу – отдельно для реляционных баз данных, для XML-документов, для текстов HTML и RTF. Поскольку полнотекстовый поиск в документах сравнительно дорогая операция, он выполняется один раз после загрузки документов коллекции в процессе индексации документов, которая обеспечивает нахождение в текстах терминов соответствующей онтологии и формирование индекса.

Специальный сервис предусмотрен для выявления взаимодействий между объектами, которые фиксируются путем автоматического преобразования перекрестных ссылок в гиперссылки между документами. Для этого выделяется раздел онтологии, включающий, например, список названий «экологических отношений» между животными в экосистемах, то есть ролей (враги, конкуренты, паразиты, болезни и т.п.) и их значений (списки врагов, конкурентов, паразитов и т.п.). В процессе просмотра индексов выявляются соответствующие отношения и их значения, по ним определяется коллекция, формируется запрос к индексу и осуществляется поиск соответствующего вида животного или растения, а его логический адрес присоединяется к значению в исходной записи и т.д. до окончания просмотра всего индекса. После этого выполняется просмотр документов коллекций и по установленным логическим адресам формируются гиперссылки. Например, «враги зайца» – волк, лиса, сова и т.д., для формирования гиперссылок сначала в индексе находятся соответствующие им записи и их логические адреса присоединяются к названиям «врагов». Затем в коллекции находится документ «Заяц», в нем – рубрика «Враги» и для каждого «врага» на основе логического адреса формируется гиперссылка. Реализация данного сервиса позволит

автоматически связывать в единый комплекс набор коллекций, описывающих флору и фауну региона.

## 6 Заключение

В данной статье представлены результаты исследований и разработок, связанные с применением онтологического подхода для организации тематического поиска данных в коллекциях электронной библиотеки. Необходимость тематического поиска обусловлена желанием предоставить пользователю возможности построения «правильных» (т.е. адекватно отражающих на данный момент информационные потребности пользователя по интересующей его тематике) запросов и автоматического формирования из коллекций ЭБ релевантных запросу ответов в виде комплекса тематически связанных документов. Для решения поставленной задачи мы применили онтологический подход, суть которого в нашем случае состоит в построении словаря связей терминов (таксономии) и их определений, а также программных сервисов, определяющих правила и ограничения использования этих терминов для построения пользователем тематических запросов к коллекциям ЭБ. В перспективе планируется с помощью специалистов-предметников, как расширять словарь связей терминов и их определений для всех коллекций ЭБ, так и уточнять правила и ограничения использования этих терминов при построении тематических запросов с целью совершенствования программных сервисов формирования и исполнения таких запросов к коллекциям ЭБ. Исследовательский прототип системы основных сервисов, реализующий вышеописанный онтологический подход к организации тематического поиска данных в коллекциях ЭБ, представлен на сайте – <http://dl.krc.karelia.ru/ontology/>.

## Литература

- [1] Gruber T. A translation approach to portable ontology specification. In. Knowledge Acquisition. Volume 5. (1993) 199-220.
- [2] Leonid Kalinichenko, Michele Missikoff, Federika Schiappelli, Nikolay Skvortsov Ontological Modeling // Proceedings of the Fifth National Russian Research Conference "Digital Libraries: Advanced Methods and Technologies, Digital Collections" Saint-Petersburg, Russia, October 19-31, 2003. pp. 7–13.
- [3] Uschold M., Gruninger M. Ontologies: Principles, Methods and Applications // Knowledge Engineering Review. Volume 11. (1996).
- [4] В.Т. Вдовичин, А.Д. Сорокин, Н.Б. Луговая. Электронная библиотека научных информационных ресурсов КарНЦ РАН: состояние и перспективы развития // Труды Шестой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Пушкино, 29 сентября–1 октября 2004 г. с. 41–46.

- [5] В.Т. Вдовицын, В.А. Лебедев, С.В. Брагин, В.Г. Старкова, Н.Б. Луговая. Развитие сервисов электронной библиотеки. // Труды Всероссийской научной конференции «Научный сервис в сети Интернет: технологии параллельного программирования» (Новороссийск, 24–29 сентября 2007 г.). М., 2007. с. 305–310.
- [6] Вдовицын В.Т., Сорокин А.Д., Луговая Н.Б. Развитие программных сервисов и контента ЭБ КарНЦ РАН // Труды Седьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Ярославль, 4–6 октября 2005 года). – Ярославль: Ярославский государственный университет им. П.Г. Демидова, 2005. – с. 92–97.
- [7] Вдовицын В.Т., Лебедев В.А., Луговая Н.Б., Сорокин А.Д., Старкова В.Г. Разработка и развитие технологии публикации и поиска документов в электронных коллекциях // Труды Восьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Суздаль, Россия 17–19 октября 2006 г. – Ярославль: Ярославский государственный университет им. П.Г. Демидова, 2006. с. 162–167.
- [8] Гаврилова Т.А. Использование онтологий в системах управления знаниями // Технология менеджмента знаний  
[http://www.kmtec.ru/publications/library/.../use\\_ontology\\_in\\_suz.shtml?cmd=prin](http://www.kmtec.ru/publications/library/.../use_ontology_in_suz.shtml?cmd=prin)
- [9] Добров Б.В., Лукашевич Н.В. и др. Разработка лингвистической онтологии для автоматического индексирования текстов по естественным наукам. // Труды седьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Ярославль, 4–6 октября 2005 г. с. 70–79.
- [10] Жыжырий Е.А. Щербак С.С. Математическое обеспечение систем поиска, основанных на онтологиях // Личный сайт, посвященный Semantic Web. Shcherbak.net/mat\_ober/
- [11] Лебедев В.А., Старкова В.Г., Брагин С.В. Представление онтологии научной коллекции «Водные ресурсы региона» // Труды Шестой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Пущино, 29 сентября – 1 октября 2004 г. с. 86–92.
- [12] Лебедев В.А., Старкова В.Г., Брагин С.В. Применение онтологии для ведения и доступа к данным коллекций «Природные ресурсы региона» // Труды Седьмой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Ярославль, 4–6 октября 2005 г.). – Ярославль: Ярославский государственный университет им. П. Г. Демидова, 2005.
- [13] Овдей О.М., Проскудина Г.Ю. Обзор инструментов инженерии онтологий // Труды Шестой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Пущино, 29 сентября – 1 октября 2004 г.
- [14] Раменская М.Л. Определитель высших растений Карелии. Петрозаводск: Гос. изд-во Карельской АССР, 1960. С. 359.
- [15] Россеева О.И., Загорулько Ю.А. Организация эффективного поиска на основе онтологий // Международная конференция ДИАЛОГ. Компьютерная лингвистика и интеллектуальные технологии. <http://www.dialog-21.ru/>

## Ontology for thematic search in digital library collections

V. Vdovitsyn, V. Lebedev

The paper considers problems of constructing and applying ontology for thematic search in digital library collections. The prototype of the system of services to facilitate ontology assembly and editing, control over integrity of the collections, as well as formation of and query search for relevant documents is represented.

---

\* Данная работа выполняется при поддержке РФФИ (грант № 08-06-00085а).