

УДК 681.3

ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ КОНТЕНТА ЭЛЕКТРОННОЙ БИБЛИОТЕКИ КарНЦ РАН

В. Т. Вдовицын, В. А. Лебедев

*Институт прикладных математических исследований
Карельского научного центра РАН*

Рассмотрены пути развития информационных ресурсов электронной библиотеки КарНЦ РАН с использованием методов онтологического моделирования. При этом содержание коллекций документов согласуется с онтологией и определяется степенью изученности природных объектов Карелии. Электронная научная информация автоматически индексируется при помощи предметных онтологий. Это позволяет использовать онтологию для построения запросов контекстного поиска документов.

К л ю ч е в ы е с л о в а : онтология, термины, контент, электронная библиотека, индексация, интеграция, тематический поиск релевантных документов.

V. T. Vdovitsyn, V. A. Lebedev. ONTOLOGICAL MODELING OF KARELIAN RESEARCH CENTRE'S DIGITAL LIBRARY CONTENT

The ways to develop DL information resources by the method of ontological modeling are considered. The contents of document collections agree with the ontology and depend on the level of knowledge about the respective objects in Karelia. Scientific publications are indexed automatically by means of subject ontology. This enables application of the ontologies for building contextual search queries.

K e y w o r d s : ontology, terms, content, digital library, indexing, integration, thematic search for relevant documents.

Введение

Технология электронных библиотек (ЭБ) [Arms, 2000] позволяет создать наряду с коллекцией электронных научных публикаций сотрудников институтов КарНЦ РАН сетевой энциклопедический справочник по изученности природных и культурно-исторических объектов и систем Карелии. Содержание (контент) такого справочника будет способствовать как поддержанию и развитию научных исследований и специального образования, так и укреплению научного престижа КарНЦ РАН. Для создания

такого справочника необходимо выполнение двух следующих основных условий:

- реализация технологии ЭБ в виде полнофункционального интернет-портала,
- разработка модели создаваемого электронного ресурса (справочника), обеспечивающей актуальность, достоверность и полноту его контента, а также технологии создания контента на базе этой модели.

Технологическая основа портала ЭБ КарНЦ РАН создавалась и апробировалась на протяжении ряда лет и в настоящее время успешно

функционирует [Вдовицын и др., 2004, 2005, 2006]. Она обеспечивает выполнение следующих функций: создание электронных коллекций, последовательное наполнение их документами с одновременным формированием унифицированных метаданных для обеспечения последующего доступа к документам. При этом для каждой подключаемой коллекции может быть разработана онтология для тематического поиска, что обеспечивается соответствующей технологией [Лебедев и др., 2004, 2005; Вдовицын и др., 2007; Вдовицын, Лебедев, 2008].

Разработка моделей информационных ресурсов является одним из эффективных средств систематизации контента ЭБ. При этом модель представляет собой словарь терминов с указанием их взаимосвязей (тезаурус, таксономия, онтология), разработка которых признана сообществом экспертов приоритетной для России [Перспективные направления, 2007]. Для разработки модели справочника используется методология онтологического моделирования [Вдовицын и др., 2007; Вдовицын, Лебедев, 2008], созданная на основе обобщения опыта возникновения онтологий коллекций ряда природных ресурсов Карелии.

Онтология (Шаблон: Lang-neolat от др.-греч. $\omega\nu$ род. п. $\omega\nu\tau\omicron\varsigma$ – сущее, то, что существует и $\lambda\omicron\gamma\omicron\varsigma$ – учение, наука) – раздел философии, изучающий проблемы бытия; наука о сущем.

В информатике под онтологией подразумевается эксплицитная, т. е. явная, спецификация концептуализации, где в качестве концептуализации выступает описание множества объектов и связей между ними [Gruber, 1993; Ushold, Gruninger, 1996].

Онтология, таким образом, представляет собой попытку наиболее общего описания универсума существующего, который не ограничивался бы данными отдельных наук и, возможно, не сводился бы к ним. Можно выделить три уровня представления онтологий.

- Мета-онтология – описывает наиболее общие понятия, которые не зависят от предметных областей.
- Онтология предметной области – формальное описание предметной области, обычно применяется для того, чтобы уточнить понятия, определенные в мета-онтологии (если используется) и/или определить (зафиксировать) общую терминологическую базу предметной области.
- Онтология конкретной задачи – онтология, определяющая общую терминологическую базу задачи, проблемы.

Построение онтологий осуществляется при помощи языка описания онтологий. Самым известным из них является язык OWL [Загорюлько, 2008].

Формально онтология определяется как

$O = \langle X, R, F \rangle$, где

X – конечное множество понятий предметной области,

R – конечное множество отношений между понятиями,

F – конечное множество функций интерпретации.

Как видно, по форме – это определение графа с помеченными вершинами, где X – множество вершин, R – множество дуг, F – множество помет. Значения помет интерпретируется некоторой функцией (функциями).

Онтологию можно интерпретировать как иерархический граф связей терминов, принятых в предметных областях, с их толкованиями, пометами и функциями интерпретации помет [Фазлиев, 2008]. Основание: природные объекты, подлежащие описанию в ЭБ, обладают иерархической структурой, множество их свойств изучает комплекс научных дисциплин, иерархически соподчиненных, что и отображается в графе связей терминов.

В составе ЭБ КарНЦ РАН в первую очередь разрабатываются предметные онтологии по геологии, водным ресурсам, ботанике, зоологии, экологии, почвоведению, лесоведению и лесоводству. Они будут использоваться для:

- формирования комплекса коллекций документов по изученности природных объектов и систем Карелии;
- формирования толкований значений комплекса терминов научных дисциплин, а также их взаимосвязей;
- обеспечения автоматической индексации документов коллекций для последующего поиска;
- построения запросов на поиск релевантных документов.

Очевидно, что для реализации этих целей необходима соответствующая технология, включающая функции создания и редактирования онтологий, коллекций документов, индексирования документов, построения запросов на поиск и осуществление поиска релевантных запросу документов.

Ряд функций указанных технологий нами был осуществлен ранее и опубликован в серии докладов на российских конференциях [Лебедев

и др., 2004, 2005; Вдовицын и др., 2007; Вдовицын, Лебедев, 2008]. В работе излагаются решения и схемы технологий новых функций.

Построение онтологической модели

Онтологическая модель контента ЭБ представляет собой совокупность онтологий предметных областей (наук или научных дисциплин), объединяемых рубрикатором ГРНТИ.

Для обеспечения достоверности, актуальности и полноты предметных онтологий необходимо разработать методологию их создания, соответствующую целям ЭБ.

В качестве основы методологии принята следующая парадигма.

Изучаемые природные объекты Карелии подразделяются на классы в соответствии с классификацией наук и научных дисциплин (например, по рубрикатору ГРНТИ). Каждый класс объектов характеризуется некоторым набором свойств (атрибутов), принимающих значения из соответствующих областей (доменов). Некоторые подмножества свойств объявляются признаками и используются непосредственно или их значения для различных классификаций объектов внутри класса. Множество свойств объектов разбивается на группы (темы), изучение которых является предметом соответствующей научной дисциплины.

Каждый объект вступает во взаимодействие с другими объектами, что является основой для выделения различных систем и подсистем. В системах объекты выполняют некоторые роли (функции), которые могут иметь различные оценочные названия (враги, союзники и т. п.), или выражаются соответствующими формулами.

Каждый объект любого класса обладает некоторым строением, т. е. состоит из набора частей (объектов), вступающих во взаимодействия и является системой (агрегатом).

Выделяют внешнее строение (морфологию) и внутреннее (анатомию).

Взаимодействие объектов в системах в некотором масштабе времени может быть неиз-

менным (статика) или меняющимся (динамика). Разделение взаимодействий объектов на классы и виды определяется в соответствующих научных дисциплинах. Статика определяет устойчивость, а динамика (процессы) – внешнее поведение (этологию), внутреннее функционирование (физиологию), происхождение, становление (генетика, генезис). Термины в скобках здесь понимаются расширительно, в предметных областях они конкретизируются и детализируются.

Методология построения предметных онтологий, основанная на данной парадигме, определяет структуру графа связей понятий (точнее терминов) предметной области.

Для построения предметной онтологии необходимо установить номенклатуру (список) свойств (атрибутов) объектов класса, изучаемую данной научной дисциплиной. Затем определить их домены. Выделить свойства и/или части объекта – признаки, участвующие в классификациях, предварительно установив список классификаций и номенклатуры их классов. При необходимости зафиксировать соответствие (например, в виде табличной функции) значений признаков и классов.

Некоторые термины являются многословными сочетаниями (например, сухие сосновые и смешанные леса, сырые засфагненные луга). Такого рода термины будем трактовать как конкатенацию названий классов различных независимых классификаций. В тех же примерах: леса, луга – типы растительности; сухие, сырые – классы по влажности; сосновые, засфагненные – классы по преобладающим видам растений (эдификаторам).

Для обеспечения удобства поиска такого рода классификации в составе таксономии разнятся по уровням иерархии. В запросе они представляются в виде конъюнктивной (например, сырые \wedge засфагненные \wedge луга) или конъюнктивно-дизъюнктивной формы (например, сухие \wedge (сосновые \vee смешанные) \wedge леса (рис. 1).

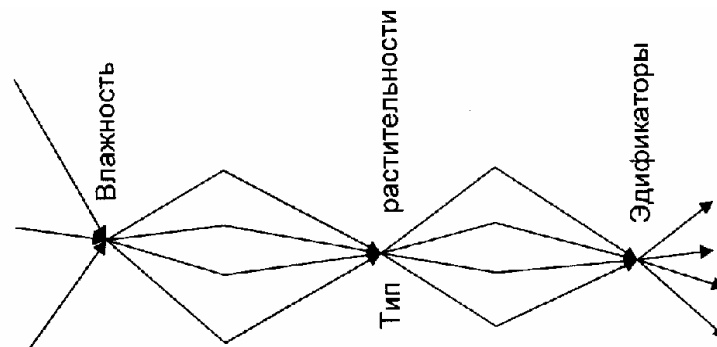


Рис. 1. Отображение классификации местообитаний по ряду независимых оснований

Далее следует устанавливать термины и названия, относящиеся к морфологии, анатомии, этологии и физиологии, т. е. зафиксировать номенклатуры названий частей объектов и систем, их функции и оценки. При этом учитываются следующие типы отношений: классификации, агрегации, синонимии и полисемии. Технология, реализующая указанную методологию, состоит в следующем:

- Корневые понятия (термины) предметных областей принимаются по рубриктору ГРНТИ.
- Начиная с корневых понятий, организуем поиск их значений (толкований) в Интернет или словарях.
- Используя найденное толкование, выделяем в нем термины более детальных понятий и ищем их толкования.
- Поступаем аналогично с терминами следующего уровня. И так до уровня значений свойств.
- В процессе поиска и нахождения терминов и их толкований фиксируем термин и URL статьи с наиболее полным толкованием его значения в связи с термином предыдущего уровня (рис. 2).

Ботаника	http://ru.wikipedia.org/wiki/Ботаника
Биомеханика	http://ru.wikipedia.org/wiki/Биомеханика
Биоценология	http://ru.wikipedia.org/wiki/Биоценология
Бриология	http://ru.wikipedia.org/wiki/Бриология
Вирусология	http://ru.wikipedia.org/wiki/Вирусология
Генетика	http://ru.wikipedia.org/wiki/Генетика
Геоботаника	http://ru.wikipedia.org/wiki/Геоботаника
Гидробиология	http://ru.wikipedia.org/wiki/Гидробиология

Рис. 2. Вид словаря терминов со ссылками на интернет-ресурсы

- Таким образом, определяем как номенклатуру терминов и их связи, так и адреса (URL) толкований. После этого материалы передаются на экспертизу специалистам-предметникам, и по результатам экспертизы итерационно выполняется построение таксономий и механизма ссылок на толкования терминов.

В целом таксономия онтологии будет иметь структуру иерархического графа (древовидного или с полуциклами), фрагмент которого представлен на рис. 3.

Вершины графа – термины, дуги – отношения между ними (классификации и агрегации), тип отношения помечается в узле разветвления. Отношения синонимии выделены в отдельную структуру (словарь). Полисемия (т. е. наличие

одинаковых по написанию терминов) разрешается ввиду того, что такие термины могут находиться только в разных частях структуры (являются терминами разных наук).

Реализация таксономии представляется в виде таблицы (рис. 4), точнее базы данных (реляционной или объектной). Технология загрузки и редактирования таксономии и словаря синонимов отработана [Загорулько, 2008]. На рис. 5 показан пример занесения нового термина.

Технология формирования контента ЭБ

Создание онтологических моделей дает возможность разработки системного плана развития контента ЭБ и его реализации. При этом материалы контента подразделяются на следующие категории:

- толкования терминологии;
- состояния изученности объектов и систем;
- проблемы, мероприятия, обсуждения;
- коллекции текущих публикаций.

Как указано выше, толкования терминов могут быть включены в состав ЭБ посредством ссылок на соответствующие ресурсы Интернет или словари.

Данные об изученности природных объектов и систем Карелии в основном могут быть подготовлены сотрудниками КарНЦ РАН. Однако ряд коллекций может быть импортирован из Интернет, например, такие коллекции как: минералы, горные породы, химические элементы и соединения, определение видов растений и др. Также могут быть задействованы сайты заповедников, природных парков, музеев и др.

Интеграция «сторонних» материалов состоит в том, чтобы обеспечить доступ к ним посредством поисковых сервисов, имеющихся в ЭБ. Это сервисы поиска: по названиям коллекций и их документов и с использованием онтологий (для формирования тематических запросов).

Каждая коллекция (собственная или включаемая) должна пройти процесс импортирования, который включает формирование: записи в списке коллекций и списка документов коллекции.

Записи в списке коллекций содержат название коллекции и ее URL в виде гиперссылки. Аналогично записи в списке документов также содержат их названия и гиперссылку на текст документа. Для формирования этих списков имеется соответствующая технология [Вдовичин и др., 2004, 2005, 2006].

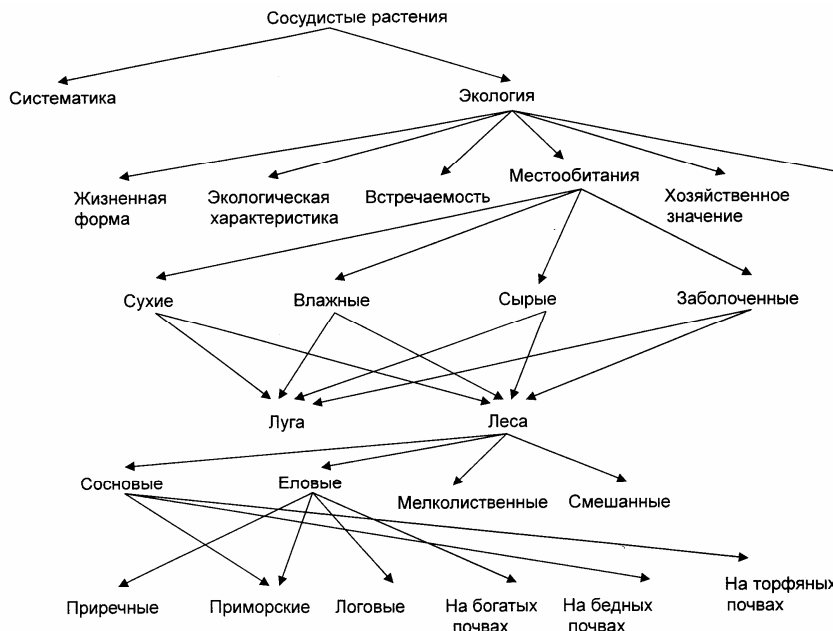


Рис. 3. Фрагмент структуры онтологии, включающий классификации терминов по независимым основаниям

Метка	Предок	Потомок
Биология А	Экология	Сообщества (экосистемы, биоценозы)
		Связи
		Виды
		Популяции
		Охрана окружающей среды
	Сообщества	Суша
		Пресные воды
		Моря
Сообщества суши К	Типы	Биосфера
		Зона
		Подзона
		Район
		Ландшафт
		Биогеоценоз
		Местообитание (биотоп)
Сообщества суши К	Зона	Арктическая пустыня
		Тундра
		Лесотундра
		Лес
Сообщества суши К	Подзона	Северная тайга
		Средняя тайга
		Южная тайга

Рис. 4. Фрагмент онтологии по биологии (начало структуры по экологии)

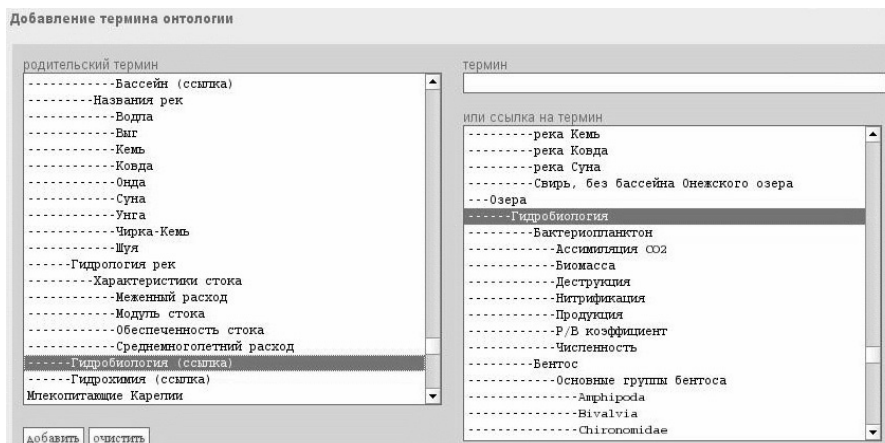


Рис. 5. Пример добавления терминов в состав онтологии по водной экологии

Отличие процесса регистрации привлекаемых коллекций заключается в том, что их документы могут быть представлены в различных форматах (HTML, PDF и др.) и могут не содержать списка документов в явном виде. Таким образом, необходимо будет разработать дополнительные технологические средства для формирования списков документов привлекаемых коллекций, аналогично тому, как формируются списки терминов онтологий и их толкований.

Для обеспечения тематического поиска документов в коллекциях производится их индексация с использованием соответствующей предметной онтологии. Структура индексного файла – это таблица, которая содержит имя документа, его URL и список встречающихся в его тексте терминов в порядке их иерархии и связей в онтологии.

Документы ЭБ по степени структуризации можно разделить на три категории: базы данных (таблицы), слабоструктурированные (XML-документы) и неструктурированные (статьи в форматах PDF, HTML и т. п.). Таблицы и XML-документы структурно соответствуют структуре онтологии, поэтому процесс их индексации сравнительно прост. Документ прочитывается «поэлементно», индексируется только текст, который находится между тегами элементов XML-документа, для которых определено соответствие с рубриками онтологии. Соответствие рубрик документа рубрикам онтологий обеспечивает сохранение в индексе порядка терминов, принятого в онтологии. Это важно для организации поиска релевантных запросу документов.

Неструктурированные документы могут содержать термины не в порядке их подчиненности в онтологии. Тогда, если не принять особых мер при их индексации, индекс документа будет содержать список терминов в порядке их нахождения в тексте, а не в порядке, принятом в онтологии, что впоследствии будет порождать «информационный шум».

Принята следующая схема индексации документов. Во-первых, в структуре графа таксономии выделен дополнительный уровень для выделения подграфов, близких по составу соответствующей коллекции документов ЭБ. Во-вторых, множество публикаций сотрудников КарНЦ РАН разделено на части (названные коллекциями), содержание документов которых близко по смыслу к соответствующей рубрике ГРНТИ. При этом статьи комплексных сборников (например, юбилейный сборник «Северная

Европа в XXI веке: природа, культура, экономика») разделяются между разными коллекциями. Это решение значительно сокращает непроизводительные затраты времени на установление соответствия между содержанием статьи и требуемым фрагментом онтологии.

Далее, в тексте документа в процессе его чтения сначала ищутся термины, близкие к корню онтологии (темы, подтемы). И если найден один, то дальше ищутся термины, подчиненные ему вплоть до «листьевых» терминов, и они помещаются в индекс. Далее ищется следующий термин корневых уровней и подчиненные ему и т. д. В результате список терминов в индексе будет иметь порядок, соответствующий онтологии.

Процесс индексации документов привлекаемых коллекций должен предусмотреть несколько ветвей предварительной обработки форматов документов, например, перевод в текстовый формат для осуществления единообразного просмотра текстов в процессе индексации. Индексные файлы образуют для коллекций, относимых к соответствующей дисциплине.

После выполнения указанных операций сторонние коллекции считаются интегрированными в ЭБ, и доступ к ним осуществляется при помощи сервисов нашей ЭБ.

Интеграция статей толкований терминов онтологий отличается тем, что списки статей, относящихся к данной предметной онтологии, включаются в соответствующий индексный файл. Причем индексация статьи может не производиться.

Реальные объекты, описываемые в документах коллекций, вступают между собой в различные отношения, которые указываются в виде их ролей в составе системы. Целесообразно использовать эту информацию для создания гиперссылок между документами. В результате получаем не просто наборы документов, а комплексы связанных документов, что полезно при их изучении. Для решения этой задачи разработана соответствующая технология [Вдовицын, Лебедев, 2008]. В итоге получим распределенную библиотеку, содержащую описание классов природных объектов Карелии и толкования терминов онтологии.

Поиск в ЭБ релевантных документов

Преимущество в использовании онтологий для формирования запросов на поиск заключается в том, что запрос в этом случае представляет

собой фрагмент таксономии, в котором термины связаны в иерархию. Тем самым запрос уже не является простым списком терминов, а отражает их зависимость. При этом устраняется возможная полисемия терминов и тем самым отсекается значительная часть «информационного шума» в отклике на запрос.

Ранее нами была предложена обобщенная схема запросов [Вдовицын, Лебедев, 2008], которая представляет собой редукцию предикатного выражения, а именно, нетерминалы в угловых скобках обозначают предикаты вида $X = a$, где X – слово в составе индекса, a – термин в запросе. С учетом объединения предметных онтологий на основе рубрикатора ГРНТИ схема запроса принимает следующий вид:

$$\{ \langle \text{Рубрика ГРНТИ} \rangle \langle \text{коллекция} \rangle \} [\langle \text{класс} \rangle \wedge \vee \dots \wedge] [\langle \text{агрегат (тема)} \rangle \wedge \vee \dots \wedge] [\langle \text{характеристика} \rangle \wedge] [\langle \text{список значений} \rangle \wedge \vee] \dots \wedge \vee$$

$$[\langle \text{характеристика M} \rangle \wedge] [\langle \text{список значений} \rangle \wedge \vee] [\langle \text{тема} \rangle \wedge \vee \dots \wedge]$$

$$[\langle \text{характеристика N} \rangle \wedge] [\langle \text{список значений} \rangle \wedge \vee] \dots \wedge \vee$$

$$[\langle \text{характеристика N + K} \rangle \wedge] [\langle \text{список значений} \rangle \wedge \vee]$$

$$[\langle \text{класс} \rangle \wedge \vee \dots \wedge] [\langle \text{тема} \rangle \wedge \vee \dots \wedge] [\langle \text{характеристика} \rangle \wedge]$$

$$[\langle \text{список значений} \rangle \wedge \vee],$$

где
 $\langle \text{список значений} \rangle := \langle \text{значение} \rangle$
 $\wedge \vee \langle \text{список значений} \rangle$.

Эта схема отображает возможность построения максимально точного для данной онтологии запроса на поиск. Наличие квадратных скобок указывает на возможность формирования и самых простых запросов, состоящих из одного термина. При использовании дизъюнкций в составе фрагментов, заключенных в квадратные скобки, необходимо правильно расставить круглые скобки, чтобы учитывать приоритеты логических операций.

Нетерминалы «Рубрика ГРНТИ» и «коллекция», заключенные в фигурные скобки, определяют как раздел онтологии, так и коллекцию документов, в которой должен выполняться поиск. Очевидно, что поиск осуществляется в одной коллекции. При необходимости поиска в большем их числе запросы должны повторяться.

Нетерминалы «класс», «тема», «характеристика», «значение» отражают иерархическую

структуру онтологии. При этом нетерминалы «класс» и «тема» подразумевают возможность иерархических классификаций.

Пример запроса:

Рубрика ГРНТИ	Водное хозяйство \wedge
Коллекция	Озера \wedge
Тема	Гидробиология \wedge
Подтема	Фитопланктон \wedge
Характеристика	Биомасса \wedge Численность \wedge Сезонное изменение биомассы

Очевидно, что построение запроса по указанной схеме непростая задача, поэтому предусмотрены средства оказания помощи пользователю в составлении запроса. Сначала он в процессе поиска в глубину по онтологии формирует список необходимых терминов, и только далее, обдумав свои потребности, переносит термины в запрос в порядке, определяемом иерархией и соответствующей инструкцией. При этом расстановка знаков конъюнкции и дизъюнкции и формирование оператора Select выполняется при помощи программного сервиса, который контролирует допустимость конъюнктивных связей между терминами, как это показано ниже.

Следует обратить внимание на двойственность семантики союза «и» в русских текстах. Часто его семантика соответствует не конъюнкции, а объединяющему «или», т. е. дизъюнкции. Например, выражение «сосновые и смешанные леса» должно быть представлено следующей формулой:

$$(\text{сосновые} \vee \text{смешанные}) \wedge \text{леса}$$

В нашем случае онтология представляет собой множество терминов предметной области, связанных между собой отношениями классификации, агрегации и синонимии.

Классификации разбивают некоторые исходные множества на группу непересекающихся подмножеств (классов) по определенным основаниям, в качестве которых могут использоваться наличие или отсутствие у объекта определенных атрибутов (признаков) и/или определенных значений атрибутов. Классификации могут быть одноуровневыми или многоуровневыми (иерархическими), многоуровневость производит последовательная классификация сначала исход-

ного множества, а затем его подмножеств, подмножеств этих подмножеств и т. д. Одно и то же множество может быть классифицировано несколько раз с использованием различных оснований. В этом случае классификации также целесообразно разнести по уровням иерархии, иначе все разбиения попадут в один список, разобраться в котором пользователю будет затруднительно и это будет источник построения недопустимых (пустых) конъюнкций (см. рис. 1).

Таким образом, по определению в классификациях допускаются конъюнкции между терминами, лежащими на одном пути в графе онтологии. Все остальные конъюнкции являются пустыми, так как связывают непересекающиеся подмножества. В этих условиях, чтобы проверить допустимость любой конъюнкции, заданной в запросе, достаточно проверить, лежат ли входящие в нее термины на одном пути в онтологии и, если не лежат, сообщить об этом пользователю, чтобы он исключил эти конъюнкции из запроса.

Часто бывает, что классификации верхних уровней (т. е. отражающие их термины) могут отсутствовать в конкретных документах. Аналогичное бывает и в случаях множественной классификации набора документов по различным основаниям. В этом случае допустимые конъюнкции будут «пустыми», исправить положение может только сам пользователь, заменив конъюнкции дизъюнкциями.

Агрегации в отличие от классификаций позволяют представить класс объектов в виде совокупности частей или свойств. Отдельные объекты класса описываются указанием значений свойств или собственных имен частей. При поиске объектов класса в коллекциях в этих случаях допускаются конъюнкции между названиями свойств или частей. При поиске конкретных объектов следует указывать также значения свойств или имена частей. Запись конъюнкции должна быть аналогичной предыдущей схеме.

В некоторых случаях допустимые конструкции в запросах для классификаций и агрегаций вступают в противоречие. Для их разрешения достаточно пометить в онтологии классификации и агрегации различными знаками (например, классификации метятся буквой К, а агрегации – буквой А) (см. рис. 4). Тогда упомянутый выше контроль допустимости конъюнкций достаточно дополнить анализатором этих пометок.

Синонимические гнезда терминов в онтологиях представляются отдельными словарями синонимов. Когда пользователь указывает в запросе

се очередной термин, выполняется поиск в словаре синонимов, и если они есть, то автоматически в запрос помещается дизъюнкция всего синонимического гнезда. Тем самым осуществляется расширение запроса.

Заключение

Предложены технологии развития контента ЭБ посредством интеграции коллекций, описывающих природные объекты Карелии с использованием онтологических моделей предметных областей в качестве основы для обеспечения полноты контента. Для построения онтологий предложены технологии поиска в Интернет терминов предметных областей, их связей и толкований и включения их в состав онтологий. Онтологии также используются для индексации документов коллекций с целью сокращения времени на поиск. Индексные файлы, помимо индексов документов, содержат URL документов привлекаемых коллекций сторонних сайтов и статей толкований терминов. Для тематического поиска документов в коллекциях разработана технология построения запросов с использованием онтологий и технология поиска в индексных файлах.

Значительная часть указанных технологий уже разработана.

Авторы выражают благодарность В. Г. Старковой и Н. Б. Луговой за реализацию предлагаемых технологий и сопровождение ЭБ, а также фонду РФФИ, поддержавшему работу грантами № 08-07-00085а, № 09-07-12074 офи_м.

Литература

Вдовицын В. Т., Сорокин А. Д., Луговая Н. Б. Электронная библиотека научных информационных ресурсов КарНЦ РАН // Труды шестой Всерос. науч. конф. по электронным библиотекам. Пушино, 2004. С. 41–46.

Вдовицын В. Т., Сорокин А. Д., Луговая Н. Б. Развитие программных сервисов и контента ЭБ КарНЦ РАН // Труды седьмой Всерос. науч. конф. по электронным библиотекам. Ярославль, 2005. С. 92–97.

Вдовицын В. Т., Лебедев В. А., Луговая Н. Б. и др. Развитие и разработка технологии публикации и поиска документов в электронных коллекциях // Труды восьмой Всерос. науч. конф. по электронным библиотекам. Суздаль, 2006. С. 162–167.

Вдовицын В. Т., Лебедев В. А., Брагин С. В. и др. Развитие сервисов электронной библиотеки научных информационных ресурсов // Труды Всерос. науч. конф. Научный сервис в сети Интернет: технологии

параллельного программирования (Новороссийск, 24 – 29 сентября 2007). Издательство Московского университета, 2007. С. 305–310.

Вдовицын В. Т., Лебедев В. А. Онтологии для тематического поиска данных в коллекциях электронной библиотеки // Труды десятой Всерос. науч. конф. Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Дубна, 2008. С. 63–69.

Загоруйко Ю. А. Методы и технологии разработки, сопровождения и реинжиниринга онтологий // Онтологическое моделирование. Труды симпозиума. Звенигород, 2008. С. 167–200.

Лебедев В. А., Старкова В. Г., Брагин С. В. Представление онтологии научной коллекции «Водные ресурсы региона» // Труды шестой Всерос. конф. по электронным библиотекам. Пущино, 2004. С. 86–92.

Лебедев В. А., Старкова В. Г., Брагин С. В. Применение онтологии для ведения и доступа к данным

коллекции «Природные ресурсы региона» // Труды седьмой Всерос. конф. по электронным библиотекам». Ярославль, 2005. С. 87–91.

Перспективные направления развития российской отрасли информационно-телекоммуникационных технологий (Долгосрочный технологический прогноз Российской ИТ — Foresight). М., 2007. 223 с.

Фазлиев А. З. Рассуждения о понятии «онтология» // Онтологическое моделирование. Труды симпозиума. Звенигород, 2008. С. 278–296.

Arms W. Digital Library. The MIT Press. 2000.

Arms W. Automated Digital Libraries. D-Lib Magazine. V. 6. N 7–8.

Gruber T. R. A Translation Approach to Portable Ontology specification // Knowledge Acquisition. 1993. N 5.

Uschold M., Gruninger M. Ontologies: Principles, Methods and Applications // Knowledge Engineering Review. 1996. N 11.

СВЕДЕНИЯ ОБ АВТОРАХ:

Вдовицын Владимир Трофимович

руководитель лаб. информационных компьютерных технологий, к. ф.-м. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: vdov@krc.karelia.ru
тел.: (8142) 769679

Лебедев Виктор Алексеевич

старший научный сотрудник, к. э. н.
Институт прикладных математических исследований КарНЦ РАН
ул. Пушкинская, 11, Петрозаводск, Республика Карелия, Россия, 185910
эл. почта: V1777@krc.karelia.ru
тел.: (8142) 766312

Vdovitsyn, Vladimir

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: vdov@krc.karelia.ru
tel.: (8142) 769679

Lebedev, Viktor

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Science
11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia
e-mail: V1777@krc.karelia.ru
tel.: (8142) 766312