

Онтология по естественным наукам и технологиям ОЕНТ: структура, состав и современное состояние

Добров Б.В., Лукашевич Н.В.

Аннотация

На основе разумного сочетания трех подходов - традиционных информационно-поисковых тезаурусов, формальных онтологий, ресурсов типа WordNet, разработана лингвистическая онтология по естественным наукам и технологиям ОЕНТ, предназначенной для работы при автоматической обработке текста в разных приложениях обработки информации. ОЕНТ включает в настоящее время более 50 тысяч понятий, более 150 тысяч текстовых входов, 200 тысяч прямых и более двух миллионов наследуемых отношений между понятиями. В статье мы описали структурные особенности ОЕНТ. Широта предметной области и назначение онтологии определяют ряд решений, принятых при разработке структурной организации онтологии.

Ключевые слова: онтология, лингвистическая онтология, Онтология по естественным наукам и технологиям ОЕНТ, структурные особенности ОЕНТ.

Введение

Существует большое число задач поиска информации, которые удовлетворяются с использованием различных методов. Достаточно важное место занимает поиск информации для обеспечения информационных потребностей при подготовке аналитических, исследовательских документов. Для профессионального, в том числе научно-технического, поиска информации часто недостаточно наиболее распространенного контекстного поиска по словам запроса, но требуется обеспечение поиска, основанного на знаниях, - использование синонимов, возможности автоматического расширения запроса, возможностей автоматического анализа результатов запроса и помощь в интерактивном поиске.

Традиционными средствами тематического поиска научной информации в течение многих лет являлись информационно-поисковые тезаурусы. Однако такие тезаурусы создавались для их использования в процессе ручного индексирования и поиска, и даже лучшие образцы таких тезаурусов трудно использовать в процессах автоматической обработки текстов в рамках современных приложений информационного поиска. Кроме того, отношения между терминами, используемые в традиционных информационно-поисковых тезаурусах, считаются недостаточно формализованными, субъективными. Одним из способов преодоления проблемы использования традиционных информационно-поисковых тезаурусов при автоматической обработке текстов является разработка лингвистических ресурсов нового типа, более формализованных.

Отметим, что создание формализованных онтологических ресурсов в сфере естественных наук связано с рядом проблем.

Во-первых, такие ресурсы должны быть достаточно большой величины, включая десятки тысяч понятий, что затрудняет возможность их формальных описаний.

Во-вторых, формализация ограничивается развивающейся природой науки, что проявляется в существовании различных теорий, частичным и изменяющимся со временем пониманием введенных понятий.

В-третьих, на что указывают авторы работы [1], такая проблема, как гипотетическая природа онтологий. В логических онтологиях классификационная схема существует до описания конкретных явлений, в то время как в научных онтологиях классификационная схема должна наилучшим образом объяснить наблюдаемые явления. Нахождение лучших единых классификационных схем - само по себе является важнейшим научным результатом, помогающим объяснить и описать явления.

Наконец, в научных предметных областях понятия неразрывно связаны с терминами - их языковыми представителями.

Еще одной проблемой является необходимость создания большого работоспособного ресурса силами достаточно небольшого коллектива при весьма ограниченных ресурсах.

В 2004 г. были начаты работы по разработке Онтологии по естественным наукам и технологиям ОЕНТ [2, 3], предназначенной для поддержки автоматической обработки больших массивов научных документов. Широта выбранной области, сочетание разных наук связано с тем, что для конкретных разделов той или иной естественной науки необходимы знания из разных разделов этой же науки или других наук, а также математики. Действительно, значимой проблемой при структуризации знания в пределах одной области науки является трудность в отграничении данной области от других, либо исследующих те же объекты, либо применяющих аналогичные подходы. С другой стороны, доступ к знанию таких родственных научных подходов был бы крайне интересен каждому исследователю.

В настоящее время онтология ОЕНТ включает 55 тысяч понятий, 140 тысяч терминов из таких областей, как математика, физика, химия, геология, экология, биология. Между понятиями установлено более 200 тысяч прямых отношений, что в соответствии с алгеброй сочетания отношений позволяет установить суммарно более двух миллионов отношений. Описанные термины в значительной мере покрывают терминологию этих областей, вводимую в средней школе и на начальных курсах ВУЗов.

В отличие от онтологий, предназначенных для решения конкретных задач в предметных областях [4, 5], основным назначением онтологии ОЕНТ является обеспечение концептуальной базы для широкого круга задач поиска информации в научной сфере: сбор и формализация отношений между терминами предметной области, поддержка автоматической обработки научных текстов такой, как поиск похожих документов, автоматическая рубрикация документов, автоматическое аннотирование одного или многих документов и др.

В данной статье мы рассмотрим особенности строения, состав и современное состояние онтологии ОЕНТ. В первом разделе мы изложим

требования к лингвистическим ресурсам, предназначенным для использования в автоматическом режиме в приложениях информационного поиска. В следующем разделе описываются этапы создания ОЭНТ. В основном третьем разделе мы подробно опишем структуру лингвистической онтологии – принцип выбора единиц, используемый набор отношений и набор принципов и правил установления отношений между единицами ОЭНТ.

1. Основные принципы разработки лингвистико-онтологических ресурсов для приложений информационного поиска

Современные приложения информационного поиска работают в широких предметных областях. Поэтому лингвистические и терминологические ресурсы, создаваемые для использования в приложениях информационного поиска, должны иметь очень широкое покрытие используемой лексики и также иметь возможность применяться в автоматических режимах обработки документов и запросов.

Традиционные информационно-поисковые тезаурусы [6, 7] создавались как инструмент для помощи человеку, их структура направлена на предоставление удобств специалисту-индексатору (удаление слишком конкретных терминов, удаление близких по смыслу терминов, добавление комментариев по употреблению тех или иных дескрипторов). В связи с этим при использовании традиционных информационно-поисковых тезаурусов в автоматической обработке текстовой информации возникают существенные проблемы [8].

Другим видом информационных ресурсов являются формальные онтологии [9], одним из провозглашаемых принципов которых является независимость от конкретного языка, что затрудняет их использование в автоматической обработке текстов для приложений информационного поиска, поскольку для этого единицы формальной онтологии необходимо связать с единицами конкретного естественного языка. Кроме того, стремление к четкой формализации отношений между понятиями формальной онтологии чрезвычайно трудно соблюсти в ситуации, когда необходимо создавать сверхбольшие ресурсы, и, кроме того, приводит к проблемам при установлении связей «понятие – языковое выражение» из-за естественной многозначности многих языковых выражений.

Еще одним типом информационных ресурсов являются ресурсы типа WordNet, которые создаются для описания лексики языка в соответствии с лингвистическими традициями [10]. Анализируя попытки создать терминологические ресурсы на основе WordNet, следует отметить, что структура WordNet не приспособлена для описания терминологии. Раздельное описание частей речи, слишком большой набор несвязанных между собой значений, недостаточная проработанность принципов включения многословных выражений, – все это приводит к проблемам разработки и использования терминологических ресурсов, созданных на базе модели WordNet.

Вместе с тем, в каждом из рассмотренных типов ресурсов есть качества, которые должны присутствовать в большом лингвистическом ресурсе для информационно-поисковых приложений.

Таким образом, мы считаем, что ресурс для автоматической обработки текстов в информационно-поисковых приложениях в широких предметных областях должен сочетать принципы различных традиций и методологий:

- методологии разработки традиционных информационно-поисковых тезаурусов;
- методологии разработки лингвистических ресурсов типа WordNet (Принстонский университет);
- методологии созданий формальных онтологий.

Поясним необходимость использования этих методологий и их особенности подробнее.

Поскольку важно уметь описывать терминологию широких предметных областей, то необходимо использовать опыт разработки информационно-поисковых тезаурусов, а именно следующие характеристики:

- информационно-поисковый контекст;
- принцип выбора единиц ресурса на основе значений терминов;
- описание большого числа многословных выражений, принципы включения (невключения) многословных единиц;
- небольшой набор отношений между понятийными единицами.

Так как предполагается использовать лингвистический ресурс в автоматическом режиме обработки текстов, то необходимо использовать методологию разработки лексических ресурсов типа WordNet, в которой важны следующие положения:

- понятийные единицы создаются на основе значений реально существующих языковых выражений;
- многоступенчатое иерархическое построение лексико-терминологической системы понятий;
- принципы описания значений многозначных слов и выражений.

Из методологии разработки формальных онтологий важны следующие положения:

- разработка лингвистической онтологии как иерархической системы понятий;
- использование для описания отношений формально определяемых отношений с формальными свойствами;
- в качестве аксиом (правил вывода) использование свойств транзитивности и наследования таксономических отношений и транзитивности отношений онтологической зависимости.

Именно эти принципы положены в основу разработки нескольких больших ресурсов для информационного поиска: Общественно-политического тезауруса, Тезауруса русского языка РуТез [8, 11] и ряда других.

Вышеперечисленные ресурсы имеют одинаковую структуру. Они являются онтологиями, поскольку описывают понятия внешнего мира и отношения между ними, которые устанавливаются в соответствии с требованием правомочности расширения запроса по иерархии связей при информационном поиске.

Эти ресурсы принадлежат к особому классу онтологий, так называемым лингвистическим онтологиям, поскольку введение понятий в значительной мере мотивируется значениями языковых единиц, относящихся к предметной области ресурса. В то же время они являются тезаурусами, поскольку каждое понятие связано с набором языковых выражений (слов, терминов, словосочетаний), которыми это понятие может быть выражено в тексте, – такой набор текстовых входов понятий необходим для использования онтологий для автоматической обработки текстов.

На основе сочетания перечисленных принципов создается и онтология ОЕНТ.

2. Этапы разработки онтологии ОЕНТ

Основной задачей при создании лингвистической онтологии большого размера силами небольшого коллектива является максимальное использование методов автоматизации, а также фрагментов ранее созданных лингвистических онтологий. Процедура формирования первой версии онтологии ОЕНТ включала интеграцию информации из нескольких разных источников.

Для каждой науки из рассматриваемого списка (математика, физика, химия, биология, геология) были сформированы коллекции документов (от 3000 до 8000 документов, от 50 до 90 Мб). Источником коллекций являлись документы, доступные в Интернет, следующих основных типов: материалы школьных уроков, рефераты, университетские лекции, материалы специализированных сайтов.

Была произведена обработка специальными процедурами автоматического извлечения терминоподобных словосочетаний, что дало возможность проверки употребимости терминов в материалах, а также нахождения терминов, входящих в состав предметной области [12]. Частотная часть выделенных слов и словосочетаний была очищена экспертами от ошибочных словосочетаний и от общеупотребительных выражений. Получившиеся списки были объединены в единый список кандидатов в термины.

Следующим этапом стало использование терминологии, включенной в тезаурус русского языка РуТез. Этот тезаурус содержит общеупотребительную лексику, лексику и терминологию нормативно-правовых актов и материалов СМИ. Поэтому имеет значимое пересечение с терминологией практически любой значимой предметной области. Была выполнена многошаговая процедура выгрузки информации из этого тезауруса [2]. Был образован новый тезаурус, в который были догружены выражения, извлеченные из коллекций.

После появления первой версии онтологии, с текстами, относящимися к рабочим предметным областям, а также со словарями предметной области, начали работать эксперты - инженеры по знаниям.

Основными целями их работы являются следующие:

- изучая конкретные языковые выражения, их словарные определения, употребление в конкретных текстах, определить, какому понятию соответствует значение данного языкового выражения. Если такое понятие уже существует, данное языковое выражение приписывается этому понятию. Для нового понятия создается отдельная единица в иерархической сети;
- для каждого понятия по текстам набирается максимально возможное число различных слов, выражений, значения которых соответствуют этому понятию - текстовых входов понятия;
- для каждого понятия проводится концептуальный анализ для выяснения его таксономических отношений и отношений онтологической зависимости.

Поскольку эти отношения являются наиболее важными для широкого круга понятий, их часто можно выявить на основе анализа определений соответствующих терминов в терминологических словарях, употреблений в текстовых контекстах, сопоставления определений и текстовых контекстов.

Как показывает практика, в связи с многократно описанными проблемами получения знания от экспертов в предметной области [13], наиболее эффективным является максимально полная разработка ресурса на основе анализа текстового корпуса. В многочисленных исследованиях подтверждено, что к описанию предметной области следует привлекать специалистов по знаниям, а также специалистов по языку, чем полностью доверить данную работу специалистам в предметной области. Во-первых, описание структуры научного знания предметной области является другим видом деятельности, чем, собственно, исследование в рамках данной науки. Во-вторых, большинство ученых действует в рамках своей научной школы, недостаточно представляя понятийный аппарат других школ. В-третьих, каждый исследователь видит недостаточную детализацию любого понятия в рамках своей науки (вполне достаточную для информационной обработки результатов исследований).

Далее созданный проект ресурса предъявляется экспертам в предметной области, которые уже достаточно легко находят в нем возможные ошибки и неточности, могут объяснить, почему им не понравилось то или иное отношение.

Следует отметить, что на этапе разработки онтологии в качестве инженеров по знаниям выступают лингвисты, которые имеют опыт работы с текстами, лексическими значениями.

3. Структура онтологии ОЕНТ

Онтология ОЕНТ - это иерархическая сеть понятий. Каждое понятие имеет имя. Для сопоставления с текстом каждое понятие снабжается набором текстовых выражений («текстовых входов», «терминов»), значения которых соответствуют данному понятию. В качестве таких текстовых входов могут выступать однословные существительные, прилагательные, глаголы, именные и глагольные группы. Количество таких текстовых входов понятий может быть достаточно велико, например, превышать 20 единиц. При вводе нового понятия делаются специальные усилия, чтобы максимально подробно перечислить его возможные текстовые входы.

Каждое понятие связывается отношениями с другими понятиями онтологии ОЕНТ. Набор отношений тезауруса специально подобран для эффективной работы в информационно-поисковых приложениях.

Особенностью онтологии ОЕНТ (как и других тезаурусов) является то, что понятия не имеют внутренней структуры в виде атрибутов (фреймовых элементов), то есть свойства понятий описываются только посредством отношений с другими понятиями.

3.1. Единицы онтологии ОЕНТ

В онтологии ОЕНТ каждое понятие должно иметь однозначное имя, которое построено на базе его текстовых входов, и должно быть понятным носителю языка. Понятие может иметь комментарий, который пишется в случае необходимости и не является частью имени понятия.

Каждое вводимое понятие должно быть снабжено списком слов и словосочетаний, с помощью которых можно сослаться в тексте на вводимое понятие - текстовых входов. В качестве таких текстовых входов могут быть отдельные слова (существительные, прилагательные, глаголы), а также именные и глагольные группы. Текстовый вход может быть многозначным (иметь другие значения), тогда он должен

быть помечен как многозначный. Для лучшего распознавания в тексте текстовые входы онтологии ОЕНТ снабжаются последовательностью нормализованных форм всех составляющих многословного выражения (мужской род, именительный падеж, единственное число).

Языковые выражения (слова, словосочетания, термины), которые были описаны как текстовые входы одного и того же понятия, становятся неразличимыми с точки зрения онтологии ОЕНТ – онтологическими синонимами.

В онтологии ОЕНТ большое значение придается работе со словосочетаниями не только как с источниками новых понятий, но и в качестве пополнения синонимических рядов. Поскольку в процессе нашей работы выяснилось, что многие слова имеют многословные синонимы, то такие синонимы специально ищутся и ими пополняются синонимические ряды текстовых выражений, связанных с понятием. Такие многословные синонимы особенно важно найти для многозначных слов, поскольку многословные синонимы уже становятся однозначными.

Понятия в онтологии ОЕНТ могут иметь достаточно большие ряды онтологических синонимов. Приведем пример синонимического ряда, включающего несколько типов синонимов для понятия *АТМОСФЕРА НЕБЕСНОГО ТЕЛА* (по алфавиту):

Атмосфера (м - многозначное),

Атмосфера небесного тела

Атмосферная оболочка

Атмосферный

Газовая оболочка небесного тела

Газообразная оболочка (м)

Газообразная оболочка небесного тела

Как видно, синонимический ряд понятия может содержать значительно количество синтаксических вариантов словосочетаний, некоторые словосочетания образуются заменой слова-компонента на синоним. Установление соответствия таких текстовых входов понятию является наиболее простым способом обнаружения понятия в тексте.

Широта предметной области онтологии ОЕНТ связана с тем, что имеется достаточно большое число многозначных терминов. В онтологии ОЕНТ существуют два основных способа представления значений многозначных терминов.

Первым способом представления многозначности является задание одного и того же текстового входа разным понятиям тезауруса (М?многозначность). Например, текстовый вход *дерево* сопоставлен трем разным понятиям: *Дерево (растение) – Дерево (материал) – дерево (граф)*, текстовый вход *белок* соответствует понятиям: *Белок (полимер) – белок яйца – белок глаза*, текстовый вход *атмосфера* соответствует понятиям *АТМОСФЕРА НЕБЕСНОГО ТЕЛА* и *АТМОСФЕРА (ЕДИНИЦА ДАВЛЕНИЯ)*.

Второй способ представления многозначности используется в тех случаях, когда слово представлено в тезаурусе в одном значении, но если известно, что оно может употребляться и в других значениях в целевых текстах, то ему ставится специальная пометка многозначности (А-многозначность).

Например, для слова *пар* только одно значение *ПАР (ГАЗООБРАЗНОЕ ВЕЩЕСТВО)* находится в онтологии ОЕНТ, но потенциально это слово может быть словоформой нетерминологического слова *пара*. В таком случае можно помещать слово *пар* как текстовый вход к соответствующему понятию с пометкой многозначности: превалирующее значение отражено, а при появлении этого слова в другом контексте соответствующее понятие выводиться не будет. Похожая ситуация возникает со словами *отражать, отражение, последовательность, аргумент*, одно из терминологических значений которых входит в состав текстовых входов онтологии ОЕНТ, а второе значение является общеупотребительным.

3.2. Отношения между понятиями в онтологии ОЕНТ

Отношения между понятиями, описываемые в онтологическом ресурсе, предназначенном для автоматической обработки текстов в рамках информационно-поисковых приложений, должны выполнять разнообразные функции.

Во-первых, эти отношения должны использоваться в классических функциях информационно-поисковых тезаурусов для расширения поискового запроса или вывода рубрики документа.

Во-вторых, отношения важны для разрешения многозначности языковых единиц, включенных в ресурс, поскольку естественным методом реализации автоматической процедуры разрешения многозначности является сопоставление контекста употребления многозначной единицы в тексте и контекста соответствующего понятия в онтологическом ресурсе.

В-третьих, отношения в онтологическом ресурсе могут использоваться для выявления лексической связности в текстах с целью применения выявленной структуры текста для улучшения качества обработки текстов.

Для реализации любой из этих функций необходимо осуществление своеобразного логического вывода: встретив вхождение некоторого понятия в тексте, нужно делать многошаговые проходы по отношениям.

В большинстве информационно-поисковых тезаурусов используется очень небольшой набор отношений между дескрипторами: отношение *выше-ниже* и отношение ассоциации. При этом отношение ассоциации часто рассматривается как наиболее проблемное отношение по следующим причинам:

- по принципам установления это отношение является симметричным, а часто обозначаемые им отношения явно не симметричны,
- это отношение часто устанавливается субъективно,
- с этим отношением возникают серьезные проблемы при использовании в автоматических режимах расширения запроса, вывода рубрики

и т.п.

Поэтому в литературе имеется много предложений по замене отношения ассоциации на более подробные наборы отношений, что было реализовано в ряде тезаурусов, например, медицинской тематики.

Кроме того, в последнее время активно обсуждается вопрос о преобразовании существующих информационно-поисковых тезаурусов в более формализованные онтологические ресурсы, с более подробной системой отношений, с возможностью логического вывода на базе аксиом, связанных с каждым отношением [14].

Однако, на наш взгляд, существуют серьезные проблемы на пути преобразования информационно-поискового тезауруса в такого рода онтологию и использование в приложениях информационного поиска, поскольку при автоматическом анализе текста далеко не всегда можно быть уверенным в том, что в тексте упомянуто именно определенное отношение между сущностями, а это значит, что сложные онтологические формализмы, построенные на шатком базисе, не смогут работать эффективно.

Таким образом, мы полагаем, что среди потенциального множества отношений понятия наиболее стабильно можно опираться на те отношения, которые не исчезают, не меняются в течение всего срока существования любого или подавляющего большинства экземпляров понятия. В условиях невозможности использования сложных правил вывода для осуществления вывода по тексту важно найти и описывать в тезаурусе такие типы отношений, которые, с одной стороны, минимально зависят от контекста упоминания понятия, с другой стороны, обладающие свойствами транзитивности и наследования, подобно таксономическим отношениям.

В результате исследований и экспериментов мы пришли к набору отношений ресурса, предназначенного для эффективной автоматической работы в информационно-поисковых приложениях.

В онтологии ОЕНТ имеется четыре основных типа отношений.

Первый тип отношений – родовидовое отношение *ниже-выше*, представляет собой отношение класс-подкласс, обладает свойствами транзитивности и наследования.

Второй тип отношений – отношение *часть-целое*. Используется не только для описания физических частей, но и для других внутренних сущностей понятия, таких как свойства или роли для ситуаций. Важным условием при установлении этого отношения является то, что понятия-части должны быть жестко связаны со своим целым, то есть каждый пример понятия-части должен в течение всего времени своего существования являться частью для понятия-целого, и не относиться к чему-либо другому.

В этих условиях удастся выполнить свойство транзитивности введенного таким образом отношения *часть-целое*, что очень важно для автоматического вывода в процессе автоматической обработки текстов.

Еще один тип отношения, называемого *несимметричной ассоциацией* $асц2 - асц1$, связывает два понятия, которые не могут быть связаны выше рассмотренными отношениями, но когда одно из которых не существовало бы без существования другого. Например, понятие *КИПЕНИЕ* требует существования понятия *ЖИДКОСТЬ*. В онтологических исследованиях такое отношение называется отношением онтологической зависимости.

Последний тип отношений – *симметричная ассоциация* – связывает, например, понятия, очень близкие по смыслу, но которые разработчики не решились соединить в одно понятие.

Отношения *выше-ниже*, *часть-целое* и *несимметричная ассоциация* являются иерархическими отношениями. Таким образом, на основе свойств иерархичности, транзитивности и наследования для каждого понятия может быть определена совокупность понятий, которые являются для него нижестоящими понятиями по иерархии – так называемое «дерево-вниз», а также может быть определена совокупность понятий, которые являются для него вышестоящими по иерархии – так называемое «дерево-вверх». Эти иерархические деревья не обязательно являются деревьями в строгом математическом смысле слова.

Рассмотрим особенности установления отношений в онтологии, предназначенной для использования в процедурах автоматической обработки текстов.

3.2.1. Установление родовидовых отношений.

Отношения *выше-ниже*, устанавливаемые в информационно-поисковых тезаурусах, не обязательно являются таксономическими отношениями в смысле онтологического моделирования. Например, в некоторых тезаурусах (AGROVOC, EUROVOC) в качестве отношений *выше-ниже* могут записываться отношения *часть-целое*.

При разработке ресурсов для автоматической обработки текста, пригодных для логического вывода, важно, чтобы отношения, называемые одинаково, обладали одинаковыми свойствами. Мы используем отношение *выше-ниже* для обозначения отношений, который обладают свойствами онтологических отношений *класс-подкласс*, а именно:

- каждый пример видового понятия в любой момент своего существования должен быть примером родового понятия,
- видовое понятие должно относиться к тому же семантическому классу, что и родовое понятие,
- видовое понятие должно наследовать основные свойства родового понятия.

Таким образом, у отношения *выше-ниже* мы предполагаем свойства несимметричности и транзитивности:

$выше(X,Y) \wedge выше(Y,Z) \rightarrow выше(X,Z)$

$ниже(X,Y) \wedge ниже(Y,Z) \rightarrow ниже(X,Z)$

$выше(X,Y) \rightarrow ниже(Y,X)$

Одной из серьезных проблем описания таксономических отношений в онтологиях является их смешение с описанием отношений «тип-роль»: от понятия-типа к понятию-роли.

Понятия-роли занимают «промежуточную» позицию между понятиями-объектами и понятиями-отношениями: роли – это то, что есть, но только в контексте того, что случается. В течение многих лет понятие роли активно обсуждается в таких областях, как концептуальное моделирование и представление знаний. Наиболее часто роль рассматривается посредством двух дополнительных понятий: «игрок» и «контекст». Например, для роли *студент* игроком является человек, а контекст определяется отношением к высшему учебному заведению [15, 16, 17].

По мнению многих исследователей ошибкой является описание понятия-роли как вышестоящего для понятия-типа (т.е. игрока) [15, 16, 17]. Например, поскольку работодателем может быть человек или организация, то понятие *работодатель* (см. рис. 1) может быть ошибочно представлено как вышестоящее, родовое понятие, а понятия *ЧЕЛОВЕК* и *ОРГАНИЗАЦИЯ* представляются как нижестоящие, видовые понятия. Однако такое представление неточно описывает свойства сущностей, поскольку не каждый человек является работодателем. Нарушается основной принцип установления таксономических отношений. Кроме того, отношение «тип-роль» очень неустойчиво в разных контекстах, что может привести к неправильному логическому выводу.



Рис.1. Расположение роли над типами сущностей нарушает основной принцип установления родовидовых отношений

Проблема смешения таксономических и ролевых отношений связана с тем, что в текстах эти отношения часто выражаются сходными языковыми конструкциями. При разработке ресурса для автоматической обработки текстов много информации приходится вводить в тезаурус на основе знаний, полученных из текстов.

Например, следующее определение кажется хорошим источником родовидовых отношений между понятием *АТМОФИЛЬНЫЕ ЭЛЕМЕНТЫ* и понятиями, соответствующими группе указанных химических элементов:

АТМОФИЛЬНЫЕ ЭЛЕМЕНТЫ (a. *atmophile elements*; н. *atmophile Elemente*; ф. *elements atmophils*; u. *elementos atmyfilos*) — группа химических элементов, выделенная на основании преобладающей роли их газообразных соединений в геохимических процессах и накопления многих из них в атмосфере (см. *Геохимическая классификация элементов*). К ним относят H, B, C, N, Cl, Br, I, He, Ne, Ar, Kr, Xe, которые либо сами в свободном состоянии являются газами (N₂, He и т.п.), либо образуют устойчивые в геологической обстановке газообразные соединения (CO₂, H₂O и др.). (*Горная энциклопедия* - <http://www.mining-enc.ru/a/atmofilnye-elementy/>).

Определение электролита:

Электролит - проводник второго рода; вещество, обладающие ионной проводимостью. Электролитами являются:

- расплавы солей, оксидов или гидроксидов;
- растворы солей, кислот или оснований в полярных растворителях;
- а также твердые электролиты.

Может показаться основанием, например, для установления отношения, что соль (как химическое соединение) является видом электролита.

Однако в таких случаях нужно помнить, что *электролит* является ролью *вещества* - вещество становится электролитом, если только оно попадает в некоторые условия. А соль как химическое соединение является типами веществ. И атмосферными веществами химические элементы становятся, только попадая в атмосферу. Устанавливая родовидовую связь от типа к роли, мы сообщаем системе некорректное знание, состоящее, например, в том, что любое вещество, относящееся к классу солей, в любой момент времени своего существования в любой ситуации является электролитом, что далеко не так.

Возникает вопрос, можно ли отразить полученную из вышеприведенных фрагментов информацию, выразив ее набором более «надежных» отношений. Мы обычно пытаемся применить несколько способов.

Во-первых, если мы предполагаем, что в нашей предметной области большинство примеров того или иного типа будут использованы в некоторой роли, то все-таки устанавливается родовидовое отношение от типа как вида к роли как роду, которое снабжается пометкой «В» - что означает «возможно по умолчанию».

Например, мы можем установить такое отношение между понятием *СОРБИНОВАЯ КИСЛОТА* и *КОНСЕРВАНТ*, если посчитаем, что это основное применение сорбиновой кислоты в нашей предметной области, и вероятность встретить в текстах обсуждение сорбиновой кислоты в других применениях (например, в органическом синтезе) в нашей области не слишком велика:

СОРБИНОВАЯ КИСЛОТА

выше_в КОНСЕРВАНТ

Однако не рекомендуется устанавливать такое отношение между понятиями *ПОВАРЕННАЯ СОЛЬ* и *КОНСЕРВАНТ*, поскольку основное применение поваренной соли совсем другое. Даже если бы мы установили такое отношение (ввели бы еще пометку для неосновных ролей), то нужно учитывать, что для автоматической системы обработки текстов невозможно качественно учитывать контекст употребления поваренной соли в тексте, чтобы разобраться, можно использовать это отношение или нет.

Таким образом, в некоторых случаях мы все-таки размещаем понятия-роли выше по иерархии, чем понятия-типы, однако отмечаем такое отношение специальной пометкой. Мы применяем это отношение только для описания знания о предметной области, которое верно по умолчанию, то есть, с одной стороны, оно может пригодиться при обработке текстов, с другой стороны, относительно редко может привести к ошибке вывода. Для каждого типа может быть описано максимум одно такое отношение, а описания многих понятий-типов не включают такие отношения, поскольку могут выступать в самых разных ролях.

На примере описания понятия *электролит* может быть продемонстрирована еще одна возможность описания отношений между ролями и типами в онтологии ОЕНТ. Можно попытаться ввести дополнительное понятие для ситуации соли в роли электролита. Если это важно для данной сферы, то это наше желание обычно поддерживается и языком предметной области – для такого понятия существует одно или более употребительных языковых выражений. И в нашем случае существует и активно употребляется такое словосочетание, как *солевой электролит*.

Таким образом, мы можем ввести понятие *СОЛЕВОЙ ЭЛЕКТРОЛИТ* и установить следующие отношения:

СОЛЕВОЙ ЭЛЕКТРОЛИТ

выше *СОЛИ*

выше *ЭЛЕКТРОЛИТЫ*

Тем самым мы корректно отражаем знание, полученное нами из прочитанного определения.

Поскольку введение дополнительных понятий может серьезно усложнять описание понятий в ресурсе, такой способ используется лишь в тех случаях, когда такие дополнительные понятия действительно используются в предметной области, как в случае понятия *СОЛЕВОЙ ЭЛЕКТРОЛИТ*.

Такая же техника может быть использована и для описания понятия *АТМОФИЛЬНЫЙ ЭЛЕМЕНТ*. Это понятие связывается с соответствующими химическими элементами не напрямую, а посредством введения ряда дополнительных понятий таких, как *АТМОСФЕРНЫЙ ВОДОРОД*, *АТМОСФЕРНЫЙ ГЕЛИЙ* и др.

Таким образом, в реальном ресурсе, создаваемом для работы в приложениях в широкой предметной области, приходится применять несколько разных подходов к описанию иерархий типов и ролей, обсуждаемых в литературе как альтернативные.

В литературе обсуждаются и более сложные представления для более адекватного описания взаимоотношений между типами и ролями, однако при создании достаточно больших онтологических ресурсов важно сохранить относительно простую схему описания. Кроме того, сложные схемы описания отношений затруднительно использовать при автоматической обработке текстов.

3.2.2. Отношение часть-целое

Отношение *часть-целое* играет существенную роль во многих предметных областях. Одним из важных свойств, которые часто постулируются у отношения *часть-целое*, является транзитивность этого отношения, на основе которой может строиться многошаговый логический вывод. В то же время многие исследователи указывают на нарушения его транзитивности [18, 19, 20, 21].

Если обсуждать наследование свойств по отношению *часть-целое* в ресурсе, предназначенном для автоматической обработки текстов в информационно-поисковых приложениях, то наиболее важной операцией, которую необходимо обеспечить, является релевантность обсуждения частей обсуждению целого. То есть необходимо описывать отношения *часть-целое* так, что если текст или его некоторый фрагмент посвящен обсуждению части, то с большой вероятностью этот текст (или его фрагмент) будет релевантен и обсуждению целого.

Здесь может быть приведено следующее возражение: если в тексте говорится о покупке деталей автомобиля, это не означает, что текст обсуждает покупку автомобиля. Мы этого и не утверждаем. Ясно, однако, что текст, обсуждающий покупку деталей автомобиля, релевантен поиску по обобщенному запросу «Автомобили».

Важным условием для обеспечения такого наследования, на наш взгляд, является зависимость существования части от существования целого. Действительно, если все существование некоторой части связано с существованием целого, то и тексты, обсуждающие эту часть, будут иметь непосредственное отношение и к целому, даже если это целое в тексте явно не упомянуто.

Этим требованием мы обеспечиваем выполнение рекомендации тезаурусных стандартов в том, что описание иерархических отношений должно быть независимо от контекста их упоминания [7]. Описание таких независимых от контекста, «надежных» отношений в ресурсах, предназначенных для автоматической обработки текстов, имеет большое значение, поскольку в автоматическом режиме часто бывает невозможно использовать контекст для подтверждения существования того или иного отношения.

Итак, описывая отношения *часть-целое* в информационно-поисковых ресурсах, предназначенных для автоматической обработки текстов, мы опираемся на три основных принципа:

1. часть должна быть зависима от целого;
2. уничтожение или изменение части влечет изменение целого;
3. свойство релевантности наследуется от части к целому: если в тексте обсуждается часть, то этот текст обсуждает и целое.

Два первых принципа заменяют в тезаурусе для автоматического индексирования правила, установленные стандартами для традиционных информационно-поисковых тезаурусов: независимость отношения от контекста и соответствие семантических типов части и целого.

Однако этим принципам соответствуют также свойства, которые зависят от своих носителей, а также роли, зависящие от своих ситуаций. В качестве примеров таких зависимых свойств можно привести следующие:

ГРУЗОПОДЪЕМНОСТЬ – *ТРАНСПОРТНОЕ СРЕДСТВО*,

ВОДОИЗМЕЩЕНИЕ – *СУДНО*,

ОКТАНОВОЕ ЧИСЛО – *МОТОРНОЕ ТОПЛИВО*.

В качестве примеров зависимых ролей можно привести следующие роли:

ДУБИТЕЛЬ – *ДУБЛЕНИЕ КОЖИ*.

Таким образом, мы описываем как части разнообразные внутренние сущности и характеристики объекта, проявляющие зависимость своего существования от существования целого объекта.

В случае, когда понятие-часть может принадлежать нескольким целым, то можно использовать несколько возможностей для описания такого отношения, которые обычно связаны с введением дополнительных понятий.

Первый способ подходит в тех случаях, когда у исходной части есть подвид, который является зависимой частью исходного целого. Так, неправильно описывать в тезаурусе, что *ДВИГАТЕЛЬ* – это часть *АВТОМОБИЛЯ*, поскольку не все двигатели являются частями автомобиля. Необходимо ввести дополнительное понятие *АВТОМОБИЛЬНЫЙ ДВИГАТЕЛЬ* как вид понятия *ДВИГАТЕЛЬ* и описать понятие *АВТОМОБИЛЬНЫЙ ДВИГАТЕЛЬ* как часть *АВТОМОБИЛЯ*.

Другим способом является введение обобщающего понятия для всех целых, к которым может принадлежать часть, и установить отношение между частью и именно этим целым.



Рис. 2. Описание взаимоотношений между понятиями альдегидная группа, альдегидокислоты, альдегидоспирты

Здесь можно привести пример из химии: альдегидная группа входит в такие соединения, как *альдегидокислоты*, *альдегидоспирты* и т.п., но имеется такое обобщающее выражение как *альдегидное соединение*. Таким образом, можно ввести понятие *АЛЬДЕГИДНОЕ СОЕДИНЕНИЕ*, описать как его виды понятия *альдегидокислоты* и *альдегидоспирты*, а понятию *альдегидная группа* установить отношение *целое* с понятием *АЛЬДЕГИДНОЕ СОЕДИНЕНИЕ* (рис. 2).

В некоторых случаях можно воспользоваться обоими методами. Так, для описания отношения *РЕАКТИВНЫЙ САМОЛЕТ – РЕАКТИВНЫЙ ДВИГАТЕЛЬ* может быть введено дополнительное понятие *АВИАЦИОННЫЙ РЕАКТИВНЫЙ ДВИГАТЕЛЬ* и/или дополнительное понятие *РЕАКТИВНЫЕ СРЕДСТВА ПЕРЕДВИЖЕНИЯ*.

При таких операциях ввода дополнительных понятий мы стараемся использовать те языковые выражения, которые реально существуют в описываемой предметной области. Если необходимого языкового выражения не существует, то может быть принято решение не описывать такое отношение часть-целое.

Имеется только одна возможность «смягчения» позиции по поводу описания зависимых частей, которая возникает в тех случаях, когда некоторая часть входит в целое в подавляющем большинстве случаев, принадлежит этому целому по умолчанию. В подобных случаях такое отношение может вводиться в тезаурус, но помечается специальной пометкой «В».

Используемый в настоящее время набор свойств отношения *часть-целое* таков:

$часть(X, Y) \leftrightarrow целое(Y, X)$

$целое(X, Y) \wedge целое(Y, Z) \rightarrow целое(X, Z)$ – транзитивность отношения

$выше(X, Y) \wedge целое(Y, Z) \rightarrow целое(X, Z)$ – наследование отношения *целое* по отношению *выше-ниже*.

3.2.3. Отношение направленной ассоциации в онтологии ОЕНТ

В традиционных информационно-поисковых тезаурусах одним из самых распространенных видов отношений являлось отношение ассоциации, которое было наиболее трудно определить. Некоторые источники излагают наиболее подробные принципы установления ассоциативных отношений, перечисляя разные семантические типы отношений, поскольку в противном случае отношения будут устанавливаться непоследовательно [22].

Американский стандарт Z39.19 [7] описывает наиболее общее правило установления ассоциативного отношения между дескрипторами таким образом, что это отношение стоит устанавливать между двумя дескрипторами, если при употреблении одного термина другой термин как бы подразумевается. Более того, один термин часто есть необходимый элемент определения другого термина, например, термин *клетка* составляет необходимую часть определения термина *цитология*.

Анализ, проведенный в работе [23] показал полезность разделения ассоциативных отношений на симметричные и несимметричные. Несимметричные отношения ассоциации наиболее близки к обсуждаемым в онтологической литературе отношениям концептуальной зависимости, относящиеся к классу отношений онтологической зависимости. Рассмотрим отношения онтологической зависимости более подробно.

3.2.3.1. Отношения онтологической зависимости

Отношение *онтологической зависимости* между сущностями А и В состоит в установлении факта зависимости существования А от существования В [24]. Это отношение известно со времен Аристотеля, который заметил, что невещественные сущности, такие, как качества и количества, зависят от вещественных сущностей.

В настоящее время это отношение стало активно использоваться при построении онтологий верхнего уровня [17, 25, 26]. В работе [27] приводятся примеры отношений онтологической зависимости в области биологии: не может быть клеточного движения без клеток, биологические процессы зависят от органов, клеток и молекул – такая зависимость является строгой.

Для выявления онтологической зависимости нужно ответить на следующий вопрос: может ли сущность (X) существовать сама по себе, или подразумевает существование чего-либо еще (Y). Так, свойство белизны зависит от вещества, например, от куска бумаги, тогда и только тогда, когда это свойство не может существовать без этого куска бумаги.

Определение. X онтологически зависит от Y тогда и только тогда, когда X существует только, если Y существует.

$D(X, Y) = \text{def}(\text{существует}(X) \rightarrow \text{существует}(Y)).$

Существует много форм онтологической зависимости.

Во-первых, можно рассмотреть онтологическую зависимость существования конкретной сущности (*экзистенциальная зависимость*), или онтологическую зависимость существования свойства сущности (*зависимость свойств*), в том числе свойство принадлежности к некоторому классу сущностей – *концептуальная зависимость* [25, 28].

Например, человек зависит от своего мозга – экзистенциальная зависимость. Свойство сущности «быть гаражом» зависит от существования автомобиля – концептуальная зависимость. Если автомобили в некотором гипотетическом мире исчезнут, то постройка останется, но ее свойство «быть гаражом» исчезнет.

Во-вторых, можно выделить *строгую зависимость (rigid)*, то есть зависимость от существования конкретной сущности, или *родовую зависимость (generic)*, то есть зависимость от существования класса сущностей. Так, человек зависит от своего мозга строгой зависимостью – мозг не может быть заменен, а от своего сердца – родовой зависимостью – сердце может быть заменено.

В-третьих, могут быть выделены *внутренняя онтологическая зависимость*, то есть зависимость от внутренних свойств или частей сущности, и *внешняя онтологическая зависимость*, то есть онтологическая зависимость от существования некоторой отдельной сущности.

В работе [28] указывается еще важность такого вида отношения онтологической зависимости как онтологическая зависимость по определению [29]:

сказать, что сущность X зависит от Y, это означает сказать, что Y необходимо (eliminably) должно быть использовано в любом определении X.

В результате авторы уточняют отношение внешней родовой зависимости следующим образом - понятие X является внешне зависимым от понятия Y, если выполняются два условия:

- определение понятия X необходимо включает понятие Y;

- если для любой сущности x, которая классифицируется как X, найдется сущность y, которая классифицируется как Y, являющаяся внешней для сущности x, то есть не являющейся частью, материалом или свойством сущности x.

3.2.3.2. Свойства несимметричной ассоциации в онтологии ОЕНТ

После многих экспериментов мы пришли к выводу, что в онтологии, предназначенной для автоматической обработки текстов, прежде всего для приложений информационного поиска, необходимо отражать *внешнюю концептуальную зависимость*, то есть зависимость существования понятия от существования другого понятия.

«Физическое» объяснение нашего выбора подтипа отношения внешней онтологической зависимости связано с тем, что в текстах широких предметных областей мы больше всего имеем дело не с конкретными сущностями, а с понятиями, концептами, классами, и, обнаружив в тексте то или иное понятие, должны уметь предположить наиболее близкие к нему понятия, которые можно использовать для разного рода логического вывода. Такими близкими понятиями как раз и являются понятия, которые связаны с исходным понятием отношением концептуальной зависимости.

Отношение внешней концептуальной зависимости является несимметричным, и мы используем для его описания отношение несимметричной ассоциации $асц1 - асц2$. Отношение $асц1$ ведет от зависимого понятия к главному понятию отношения концептуальной зависимости, а отношение $асц2$ является к нему обратным отношением.

Возникает правомерный вопрос, насколько сложно, вводя новое понятие в тезаурус, онтологию, понять, каковы должны быть отношения концептуальной зависимости, и как правильно отразить их в ресурсе.

Здесь можно рассмотреть два случая.

Во-первых, если вводимое понятие базируется на относительно свободно построенном многословном выражении, как, например, словосочетание *автомобильный завод*, то одно из слов обычно указывает на родовое понятие (*ЗАВОД*), а второе слово, в частности, может указывать и на отношение концептуальной зависимости. Действительно, понятие *АВТОМОБИЛЬНЫЙ ЗАВОД* не могло бы возникнуть, если бы не было понятия *АВТОМОБИЛЬ*.

Во-вторых, вводимое понятие может основываться на термине, который имеет определение. Здесь необходимо опереться на отношение онтологической зависимости по определению, введенное в работе [29]. Кроме того, помогает применение диагностического высказывания, подобного лингвистическим тестам [18]:

Возникновение понятия C0 зависит от существования понятия C1.

Приведем пример понятий, находящихся в концептуальной зависимости (т.е. связанных в онтологии несимметричной ассоциацией) с понятием *АТОМ*:

АТОМ	
асц2	АТОМИСТИЧЕСКАЯ ТЕОРИЯ
асц2	МОДЕЛЬ АТОМА
асц2	АТОМНО-МОЛЕКУЛЯРНАЯ СТРУКТУРА
асц2	АТОМНЫЙ СЛОЙ
асц2	АТОМНАЯ АДСОРБЦИЯ
и др.	

Как видно, посредством несимметричных ассоциаций отражаются теории и модели, связанные с атомами, структуры, которые строятся из атомов, процессы, которые происходят с атомами.

В настоящее время в приложениях используются следующие свойства отношения внешней концептуальной зависимости, обозначаемой как несимметричная ассоциация:

$асц1(X, Y) - асц2(Y, X)$

Наследование отношения несимметричной ассоциации на виды и части:

$выше(X, Y) \wedge асц1(Y, Z) \rightarrow асц1(X, Z)$

$целое(X, Y) \wedge асц1(Y, Z) \rightarrow асц1(X, Z)$

3.2.4. Отношение симметричной ассоциации в онтологии ОЕНТ

Существует несколько ситуаций, когда оправданно представление отношений между понятиями в виде симметричной ассоциации. При этом предполагается, что степень ассоциации между понятиями достаточно высокая, т. е. если два понятия C_1 и C_2 связаны отношением симметричной ассоциации, то тексты, содержащие понятие C_1 , часто релевантны запросам, содержащим понятие C_2 , и наоборот.

В онтологии ОЕНТ чаще всего симметричной ассоциацией описывается отношение между близкими по смыслу понятиями, относящимися к одному и тому же родовому понятию, текстовые входы которых используются как квазисинонимы. Например, есть близкие понятия *АВИАЦИОННАЯ МЕДИЦИНА* и *КОСМИЧЕСКАЯ МЕДИЦИНА*, также имеется множество контекстов употреблений словосочетаний *авиакосмическая медицина*, *авиационная и космическая медицина*. В некоторый момент развития тезауруса отношение между такими понятиями может быть отражено в виде симметричной ассоциации.

Используются следующие свойства отношения симметричной ассоциации:

$асц(X, Y) \rightarrow асц(Y, X)$ - симметричность отношения ассоциации

Наследование отношения ассоциации на виды и части:

$выше(X, Y) \wedge асц(Y, Z) \rightarrow асц(X, Z)$

$целое(X, Y) \wedge асц(Y, Z) \rightarrow асц(X, Z)$

Заключение

На основе разумного сочетания трех подходов – традиционных информационно-поисковых тезаурусов, формальных онтологий, ресурсов типа WordNet, разработана лингвистическая онтология по естественным наукам и технологиям ОЕНТ, предназначенной для работы при автоматической обработке текста в разных приложениях обработки информации. ОЕНТ включает в настоящее время более 50 тысяч понятий, более 150 тысяч текстовых входов, 200 тысяч прямых и более двух миллионов наследуемых отношений между понятиями.

В статье мы описали структурные особенности ОЕНТ. Широта предметной области и назначение онтологии определяют ряд решений, принятых при разработке структурной организации онтологии.

В настоящее время онтология тестируется как база знаний для автоматической рубрикации научных текстов.

Литература

1. Tsujii, J., Ananiadou, S., Thesaurus or logical ontology, which one do we need for text mining? // Language Resources and Evaluation, Springer Science and Business Media B.V., vol. 39, no 1, 2005. P. 77-90.
2. Добров, Б.В., Лукашевич, Н.В., Сеницын, М.Н., Шапкин, В.Н., Разработка лингвистической онтологии для автоматического индексирования текстов по естественным наукам // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Седьмой Всероссийской научной конференции (RCDL'2005) – Ярославль: ЯрГУ им.П.Г.Демидова, 2005. С.70-79.
3. Добров, Б.В., Лукашевич, Н.В., Вторичное использование лингвистических онтологий: изменение в структуре концептуализации // Восьмая Всероссийская научная конференция «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» (Владимир-Суздаль, 16-18 октября 2006г.). С.56-64.
4. Артемьева, И.Л., Рештаненко, Н.В., Интеллектуальная система, основанная на многоуровневой онтологии химии // Программные продукты и системы, 1, 2008. С. 84-87.
5. Привезенцев, А. И., Фазлиев, А.З., Прикладная онтология для задач молекулярной спектроскопии // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Девятой Всероссийской научной конференции (RCDL'2007) . 2007.
6. ГОСТ 7.25.-2001 Тезаурус информационно-поисковый одноязычный: Правила разработки: структура, состав и форма представления:

Межгосударственный стандарт. – Минск: Межгосударственный совет по стандартизации, метрологии и сертификации, 2001.

7. Z39.19. Guidelines for the Construction, Format and Management of Monolingual Thesauri. – NISO. 1993.

8. Добров, Б.В., Лукашевич, Н.В., Тезаурус и автоматическое концептуальное индексирование в университетской информационной системе РОССИЯ // Третья Всероссийская конференция по Электронным Библиотекам «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - Петрозаводск, С. 78-82.

9. Guarino, N., Formal Ontology and Information Systems // Proc. of the 1st International Conference on Formal Ontologies in Information Systems FOIS'98 / N. Guarino, (ed). Trento, Italy, IOS Press, 1998. pp. 3-15.

10. Miller, G., Nouns in WordNet // WordNet – An Electronic Lexical Database / Fellbaum, C (ed). – The MIT Press, 1998. pp. 23-47.

11. Лукашевич, Н.В., Добров, Б.В., Тезаурус русского языка для автоматической обработки больших текстовых коллекций // Компьютерная лингвистика и интеллектуальные технологии: Труды Международного семинара Диалог'2002 / Под ред. А.С.Нариньяни – М.: Наука, Т.2, 2002. С. 338-346.

12. Добров, Б.В., Лукашевич, Н.В., Сыромятников, С.В., Формирование базы терминологических словосочетаний по текстам предметной области // Труды пятой всероссийской научной конференции "Электронные библиотеки: Перспективные методы и технологии, электронные коллекции, 2003. С. 201-210.

13. Гаврилова, Т.А, Извлечение знаний: Лингвистический аспект // Корпоративные системы. № 10(25). С.24-28.

14. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S., Reengineering Thesauri for New Applications: the AGROVOC Example. - Article No. 257, 2004-03-17, 2004.

15. Guarino, N., Some Ontological Principles for Designing Upper Level Lexical Resources // Proc. of First International Conference on Language Resources and Evaluation. Granada, Spain, 1998.

16. Steinmann, F., The representation of roles in object-oriented and conceptual modelling // Data and Knowledge engineering. 35, 1, 2000. pp. 83-106.

17. Sowa, J., Knowledge Representation: Logical, Philosophical, and Computational Foundations. – Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

18. Cruse, D., Lexical Semantics. Cambridge. University Press. 1986.

19. Motschnig-Pitrik, R., Kaasboll, J., **Part-Whole Relationship Categories and their Application in Object-Oriented Analysis** // IEEE TSE. v. 11(5), 1999. pp. 779-797.

20. Johansson, J., On the Transitivity of the Parthood Relations // Relations and Predicates. – Frankfurt: Ontos Verlag, 2004. pp. 161-181.

21. Winston, M., Chaffin, R, Herrmann, D., A Taxonomy of Part-Whole Relations // Cognitive Science, 11, 1987. pp. 417-444.

22. Aitchinson, Y., Gilchrist, A. Thesaurus construction: a practical manual. – 2nd ed. – L.:Aslib. 1987.

23. Лукашевич, Н.В., Добров, Б.В., Отношения в онтологиях для решения задач информационного поиска в больших разнородных текстовых коллекциях // Девятая национальная конференция по искусственному интеллекту с международным участием КИИ-2004: Труды конференции. В 3-х т. - Т2. – М.: Физматлит, 2004. С. 544-551.

24. Lowe, E.J., Ontological dependence // Stanford encyclopedia of Philosophy, 2005 (<http://plato.stanford.edu/entries/dependence-ontological/>)

25. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Shneider, L., WonderWeb. Final Report. Deliverable D18. 2003.

26. Grenon, P., Spatio-temporality in Basic Formal Ontology: SNAP and SPAN, Upper-Level Ontology, and Framework for Formalization: PART I. IFOMIS Report 05/2003, Institute for Formal Ontology and Medical Information Science (IFOMIS), University of Leipzig, Leipzig, Germany. 2003.

27. Kumar, A., Smith, B. The ontology of blood pressure: a case study in creating ontological partitions in biomedicine. 2004. Способ доступа: <http://philpapers.org/rec/SMITOO>.

28. Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N., Social roles and their descriptions // Proc. of the Ninth International Conference on the Principles of Knowledge Representation and Reasoning. AAAI Press. 2004.

29. Fine, K., Ontological Dependence // Proc. of the Aristotelian Society 95, 1995. pp. 269-290.

Об авторах

Добров Б.В. - НИВЦ МГУ имени М.В. Ломоносова

dobroff@mail.cir.ru

Лукашевич Н.В. - НИВЦ МГУ имени М.В. Ломоносова

louk@mail.cir.ru