



Попова М.С.

ОБЕСПЕЧЕНИЕ МНОГОЯЗЫЧНОГО ДОСТУПА К ИНФОРМАЦИОННОМУ РЕСУРСУ ЗАДАННОЙ ПРЕДМЕТНОЙ ОБЛАСТИ И ЕЕ СИСТЕМАТИЗИРОВАННЫМ ЗНАНИЯМ

В статье рассмотрен подход по обеспечению содержательного многоязычного доступа к portalу знаний, интегрирующему знания, а также к информационным ресурсам, которые на основе онтологий относятся к заданной предметной области. Сформулированы требования к концептуальной структуре онтологий. Проанализированы направления интеграции семантических поисковых систем с онтологиями для описания предметной области поиска и требования, предоставляемые поисковыми системами к стандартам онтологических метаданных и формам их представления.

Практическая реализация данных требований обеспечит пользователей наиболее пригодными инструментами для решения стоящих перед ними задач при многоязычном доступе к информационным ресурсам заданной предметной области, а также позволит выполнять задачу распознавания различных типов информационных объектов, которые обнаруживает поисковая система.

Ключевые слова: информационные системы, онтологии, портал знаний, тезаурус, многоязычный тезаурус, информационный ресурс.

При постоянном росте объема информации, относящейся к различным отраслям знаний, обеспечивается задача эффективного информационного обеспечения научного и производственного процесса. При этом в большинстве существующих информационных систем (ИС) (например, корпоративная информационная система) и информационных ресурсов (например, интернет-каталог или портал) данные представлены, как правило, в виде текстовых документов».

Наиболее естественная форма подачи информации все же заключается в представлении ее в виде нескольких взаимосвязанных фактов. Обеспечить такое представление информации и эффективность ее поиска могут только ИС, использующие не только общие знания о мире, но и знания о предметных областях (ПрО), которые ей охватываются. На сегодняшний день эти знания представлены в виде онтологий [1].

В данной работе нами сформулированы требования к концептуальной структуре онтологий при обеспечении многоязычного доступа к информационным ресурсам предметной области, проанализированы направления интеграции семантических поисковых систем онтологий для описания предметной области поиска и требования, предоставляемые поисковыми системами к стандартам онтологических метаданных и формам их представления.

Онтологии представляют собой интероперабельное представление знаний, для которых сегодня уже созданы общепризнанные стандарты, языки представления, инструментальные средства редактирования и логического вывода, а также имеется

в наличии фундаментальная математическая база. Однако до сих пор нет четко сформулированных методов управления знаниями на основе онтологий, которые можно было бы непосредственно реализовывать в прикладных системах, в особенности – методов для распознавания различных информационных объектов (ИО), представляющих интерес для пользователя и необходимых для решения поставленных перед ним проблем. Именно восприятие и распознавание ИО – наиболее важные задачи при разработке интеллектуальных информационных систем, которые базируются на знаниях [1, 2].

ИО представляет собой некую сущность, содержащую в себе сведения о каком-либо реальном или виртуальном объекте (предмете, существе, событии, процессе и т.д.) – уникально идентифицированной материальной или нематериальной сущности реального мира и описывающую его структуру, атрибуты, ограничение целостности и, возможно, поведение.

Доступ к ИО может базироваться на онтологии соответствующей ПрО. Каждый ИО соответствует некоторому классу онтологии или является экземпляром, принадлежащим к определённому классу, а также заданной этим классом структурой. Между ИО могут существовать различные связи, семантику которых определяют отношения, заданные между соответствующими классами онтологии. Эти отношения также могут использоваться при поиске. Содержательный многоязычный доступ к систематизированным знаниям и информационным ресурсам заданной области знаний обеспечивается благодаря развитым средствам поиска и навигации [3].

Проблема восприятия, распознавания и интерпретации объектов в информационных технологиях – комплексная проблема, которая разбивается на отдельные подзадачи [4]. При этом традиционное распознавание образов, распознавание речи и распознавание текста являются лишь частными случаями значительно более общей проблемы. Распознавание предполагает обнаружение в каком-либо информационном ресурсе (ИР) сведений о том или ином ИО, интересующем пользователя. При онтологическом подходе задача распознавания может сводиться к задаче классификации, когда ИР и их фрагменты связываются с различными классами и экземплярами классов онтологии интересующей пользователя ПрО. Следует отметить, что такая классификация зависит от онтологии ПрО.

В современных интеллектуальных приложениях задача распознавания обычно формулируется следующим образом: необходимо найти:

- ИР, релевантные поставленной пользователем задаче;
- сведения, относящиеся к некоторым классам онтологии ПрО или к их экземплярам;
- структуры, которые отражены при помощи онтологии ПрО и которые являются существенными с точки зрения стоящей перед пользователем проблемы.

Для интероперабельного представления знаний сегодня все чаще применяют онтологический подход, который обеспечивает повторное и совместное использование накопленных ранее знаний.

В различных источниках предлагаются разные формальные модели представления онтологий. Во всех них присутствует:

- множество терминов (понятий, концептов), которое может подразделяться на множество классов и множество экземпляров;
- множество отношений между понятиями, в котором могут явным образом выделяться отношения «класс-подкласс», иерархические (таксономические) отношения и отношения синонимии (подобия), а также функции – специальный случай отношений, для которых n -й элемент отношения однозначно определяется $n - 1$ предшествующими элементами;
- аксиомы и функции интерпретации понятий и отношений.

Формально онтология O представляется тройкой $O = \langle X, R, F \rangle$, где X – множество концептов, R – множество отношений между концептами, F – функции интерпретации концептов из множества X и отношений из R [1]. Данная модель носит общий характер, тогда как на практике пользуются более точными моделями, в частности, связанными со стандартами и языками представления онтологий Web Ontology Language (OWL, онтологический язык для информационных сетей). Для представления и обработки OWL существует теоретический базис в виде семейства логик DL, обеспечивающий доказательность логического вывода на онтологиях, а различные средства логического вывода позволяют осуществлять вывод на структурированных данных (OWL и RDF). Проанализировав выразительные возможности разных средств представления онтологий и формальные модели онтологий, можно утверждать, что существующие технологии предлагают разные средства описания онтологий, отличающиеся возможностями и сложностью: RDF Schemas предоставляет простейший уровень для представления онтологий, а OWL Full – наиболее сложный. Выбор средства представления онтологии зависит от специфики проблемы, для которой она разрабатывается.

Онтологию можно рассматривать как основу для представления структуры информационного объекта, которую описывает класс онтологии, а различные ИР – как источники для создания экземпляров этого класса. Такой подход позволяет интегрировать информацию, поступающую из различных источников, и формировать необходимые пользователю знания. При этом задача подразделяется на несколько подзадач:

- формирование (или поиск) онтологии, отражающей структуру информационного объекта (или множества объектов), знания о которых необходимы пользователю для решения стоящей перед ним проблемы;
- поиск ИР, явно или неявно содержащих сведения об этих ПрО;

- извлечение знаний об ПрО из ИР;
- представление извлеченных знаний в форме, понятной и удобной пользователю.

Онтология может использоваться как спецификация информационной системы, поскольку она определяет знание, которое требуется для выполнения задач разрабатываемой системы. Совместное и многократное использование онтологий, относящихся к различным ПрО и приложениям, может значительно улучшить архитектуру интеллектуальных информационных систем.

Поиск информации – это процесс выявления в массиве доступной информации тех объектов, которые удовлетворяют некоторым (явно или неявно заданным) условиям. Результатом такого поиска может быть онтология, отдельный документ, фрагмент документа или же сведения о каком-либо объекте заданного пользователем типа (географическом пункте, человеке, организации, товаре и т.д.). В настоящее время информационный поиск является значительно развивающейся областью науки, чью популярность обуславливает экспоненциальный рост объема сведений. Его цель – помочь пользователю удовлетворить его информационную потребность.

Поиск информации состоит из 4-х этапов:

- определение информационной потребности пользователя и построение поискового запроса, отражающего эту потребность;
- определение совокупности доступных ИР и их характеристик;
- извлечение информации из выявленных ИР;
- предоставление пользователю результатов поиска в форме, позволяющей обеспечить удовлетворение его информационной потребности.

Существующим информационно-поисковым системам (ИПС) присущ ряд серьезных недостатков. Даже те ИПС, которые используют знания и онтологии, не всегда корректно анализируют контекст для решения проблем омонимии и синонимии в естественном языке (ЕЯ). Кроме того, недостаточно эффективно реализуется извлечение информации из найденных ресурсов, ее сопоставление с тем информационным объектом, который интересует пользователя, и интегрирование информации из различных ИР, что требует от пользователя самостоятельно выполнять эту рутинную работу. Это связано с тем, что данные операции нуждаются в знаниях о ПрО поиска, представленных в интероперабельной форме и доступных поисковой системе для повторного использования.

В определенной мере решить проблемы автоматизации создания и пополнения онтологий позволяет использование семантической разметки. На основе семантически размеченных текстов можно автоматизировать создание онтологий, содержащих термины, соответствующие тегам такой семантической разметки, извлекать информацию о связях между терминами из связей между соответствующим образом помеченными фрагментами текста. Но проблема заключается в том, что создаваемые подобным образом онтологии

используют только те теги, которые отражают информационные интересы определенного сообщества в целом, а не конкретного пользователя, решающего конкретную проблему.

Семантический поиск представляет собой разновидность автоматизированного информационного поиска, в котором учитываются смысловые аспекты запроса пользователя, доступных информационных ресурсов (ИР), среди которых осуществляется поиск, и контекста запроса [1]. При семантическом поиске его предметом может быть не конкретный ИР или его фрагмент, а информационный объект определенного класса.

Нередко системы семантического поиска сводят к поисковым системам, способным обрабатывать запросы на ЕЯ, или к системам, обрабатывающим метаданные о ресурсах. Однако семантический поиск – это более широкое понятие. Как правило, он предполагает семантический анализ естественно-языковой составляющей объектов и запроса пользователя. Для семантического анализа могут применяться контент-анализ, метод семантических падежей, ассоциативный анализ, метод тематической классификации на основе модели структурного представления текста, семантический дифференциал, латентно-семантический анализ и т.д.

Принципиальная особенность семантического поиска состоит в том, что тем или иным образом анализируются не только формальные параметры рассматриваемых объектов, но и их семантика. Эффективность поиска можно значительно повысить за счет интеллектуального анализа объектов, для которого применяются агентный и онтологический подходы. При этом онтологии могут применяться как для описания семантики контента отдельного документа и его структуры, так и для описания тех объектов, информация о которых нужна пользователю.

При переходе к семантическому поиску возникает проблема распознавания различных ИО. Распознавание ИО предполагает обнаружение в каком-либо ИР сведений о том или ином ИО, интересующем пользователя. Распознавание ИО можно рассматривать как частный случай задачи распознавания образов, которое определяется в качестве отнесения исходных данных к определенным классам через выделение существенных признаков, которые характеризуют эти данные из общей массы данных.

Распознавание образов требует проведения классификации объектов из заданного множества по имеющимся описаниям объектов и классов [1]. Стандартная постановка задачи распознавания образов заключается в том, чтобы для множества объектов M , описываемых набором признаков, и представляемого в виде объединения непересекающихся подмножеств (классов) $M = \bigcup_{j=1}^p M_j$, таких, что $\forall j \neq k, M_j \cap M_k = \emptyset$, причем для K -части объектов из M , $\forall K, K \in K, i = 1, n, K \subseteq M$ известно, к какому именно классу они относятся, требуется по набору значений признаков объектов из $L = M \setminus K$ определить класс этих объектов.

Создание интеллектуальных систем для поиска и распознавания различных информационных объектов в заданной предметной области требует разработки методов и инструментов, которые способны не только извлекать и обрабатывать новые знания, но и повторно использовать знания, полученные ранее. Одним из возможных решений этой задачи является использование онтологий, которые обеспечивают хранение, поиск, оценивание и безопасное использование онтологий, а также управление изменениями в них, персонификацией, разделением, отображением и интеграцией. При этом возникает задача формирования требований к метаописаниям онтологий и выработка единого стандарта, реализующего эти требования.

В связи с этим становится понятной необходимость исследования того, какие именно сведения об онтологиях позволяют обеспечить их использование, в частности, для задач, связанных с семантическим поиском и распознаванием разных типов информационных объектов, соответствующих классам онтологий, и какие функциональные возможности онтологий могут для этого использоваться.

Очевидно, что для семантического поиска необходимы не только методы и алгоритмы обработки семантики ИР, но и формально представленные знания о ПрО поиска, в частности, онтологии соответствующей ПрО. В общем случае можно выбрать один из трех возможных источников таких онтологий – создать новую онтологию, модифицировать уже имеющуюся или найти созданную ранее и удовлетворяющую потребностям пользователя. Для создания онтологий могут быть использованы ручное построение онтологий специалистом ПрО, автоматизированная обработка метаданных об ИР, получение онтологических знаний из естественно-языковых текстов, применение индуктивного вывода. Модификация онтологий может выполняться вручную либо автоматизированно при помощи логических операций над существующими онтологиями (пересечение, объединение, разность и т.д.).

Пользователю надо либо самому создать онтологию, отражающую его информационные интересы к определенной ПрО, чтобы потом использовать ее при поиске ИР, что довольно сложно, либо повторно использовать онтологию, созданные ранее другими исследователями и покрывающие область его интересов, или непосредственно без каких-либо изменений, или расширяя и изменяя ее. Но для этого пользователю требуются средства для поиска таких онтологий, которые связаны с нужной ему ПрО и обладают требуемой степенью сложности и детальности.

Пополнение онтологии ПрО может осуществляться при помощи многоязычного лингвистического анализа текстов, определенных пользователем в соответствии с представлениями о своих информационных потребностях. На сегодняшний день разработан ряд методов и инструментов для автоматизированного построения онтологий и тезаурусов по полнотекстовым ИР. Суще-

ствуют средства сопоставления запросов и ресурсов, ориентированные предоставление многоязычного доступа к информационным ресурсам и на семантический поиск заданной предметной области, которые могут использоваться и для поиска ресурсов других типов, а также средства сопоставления онтологий.

Таким образом, встает задача поддержки в предметной области нескольких языков. Данная задача может быть решена, путем расширения имеющихся онтологий необходимыми лингвистическими знаниями. Однако введения в онтологии дополнительных сущностей и отношений сделает ее достаточно громоздкой и непрозрачной, затруднит ее развитие, а также сопровождение. В связи с этим логичным будет решение по расширению системы знаний другим компонентом – многоязычным тезаурусом [2, 3, 4], введение которого может сделать портал приспособленным к «пониманию» разноязычных ресурсов, поддержанию навигации по его информационному пространству и восприятию запросов на разных языках.

Вопрос использования тезауруса и онтологии неоднократно обсуждался в научных работах. В основном в теории рассматривается применения тандемов «тезаурус-онтология» в задаче по обработке текстов и информационном поиске [4, 5]. Подход, представленный в данной статье, способствует обеспечению совместного использования онтологии и тезауруса не только для решения данных задач, но и для интеграции этих знаний и ресурсов, релевантных заданным ПрО в едином информационном пространстве и обеспечении эффективных навигаций по нему через использование пользователем естественного языка. В связи с этим нами предлагается использование в онтологии всех отношений, необходимых для представления знаний о данной ПрО, с оставлением чисто лингвистических отношений (синонимии, эквивалентности и т. п.) для тезауруса.

Создание предметной области обеспечивает содержательный доступ к знаниям и информационным ресурсам определенной тематики, способствует решению задач, связанных с обработкой текстовых ресурсов, представлении их содержания в виде взаимосвязанных фактов, организации поиска нужной информации и её визуализация на разных языках. Решения данных задач заключается в использовании знаний о структуре и терминологии данной ПрО, о структуре и типологии ресурсов, а также знаний о свойствах языка, на которых представляются тексты данных ресурсов.

Технологией по созданию такой ПрО предполагается организация ее систем знаний на базе интеграции многоязычных тезаурусов и онтологий связанных ПрО.

Систему знаний (KS) портала опишем как четверку вида $KS = (Os, Th, ICs, Ir)$, где Os – онтология порталов знаний; Th – многоязычные тезаурусы предметных и проблемных областей порталов знаний, ICs – информационные наполнения порталов знаний, которые строятся на основе структуры, заданной в онтологии Os ; Ir – информационный ресурс, интегрированный в информационное пространство портала знаний.

Представление онтологии требует формализма, обеспечивающего гибкие средства по описанию понятий его проблемных и предметных областей и всех разнообразных семантических связей, которые появляются между ними. Важные требования к нему заключаются в возможности выстраивания понятий ПрО в иерархию «общее-частное» и поддержке наследования свойств по этим иерархиям. Этот формализм также должен представляться как возможность заданий ограничений на значения свойств объектов ПрО и описаний семантики отношения в виде аксиом.

В качестве формализма, который удовлетворяет описанному выше, предлагается метаонтология следующего вида $O = (C, R, T, D, A, F, Ax)$, где $C = \{C_1, \dots, C_n\}$ – конечные непустые множества классов, которые описывают понятие данных предметных или проблемных областей; $R = \{R_1, \dots, R_m\}$, $R_1 \subseteq C * C$, $R = \{R_T, R_P\} \cup R_A$ – конечные непустые множества бинарных отношений, заданных на классах (понятиях); R_T – антисимметричные, транзитивные, нереклексивные бинарные отношения наследования, которые задают частичный порядок на множествах понятий C ; R_P – бинарные транзитивные отношения включений («часть-целое»); R_A – конечные множества ассоциативных отношений; $T = \{t_1, \dots, t_k\}$ – конечные непустые множества стандартных типов; $D = \{d_1, \dots, d_k\}$ – множества доменов $d_i = \{s_1, \dots, s_k\}$, где s – значения стандартного типа «string»; $TD = T \cup D$ – обобщенные типы данных, включающие множества стандартных типов и множества доменов; $A = A_C \cup A_R = \{a_1, \dots, a_w\}$ – конечные множества атрибутов, которые описывают свойства понятий C ($A_C \subseteq C * TD$) и отношения R_A ($A_R \subseteq R_A * TD$); F – множества ограничений на значения атрибутов понятий и отношений, т. е. предикатов вида $p_i = (e_1, \dots, e_m)$, где e_k – имя атрибута ($e_k \in A$), или константа ($e_k \in td_j$, где $td_j \in TD$); Ax – множества аксиом, которые определяют семантику классов и отношений онтологии. Отношения включения «часть-целое» R_P наделены свойствами транзитивности, благодаря которым в ИПС могут быть выполнены транзитивные замыкания по этим отношениям.

Набор ассоциативных отношений R_A определяет пользователь. Данные отношения позволяют организовывать ИПС, а также навигацию по контенту порталов. Важная особенность отношений R_A является то, что они имеют собственные атрибуты, которые специализируются на связи между аргументами.

Существуют ИПС, ориентированные на поиск в структурированных данных (в частности, представленных в форматах OWL и RDF). Система индексирует документы в форматах N-Triples, RDF/XML и N3 (RDF). При этом обрабатываются как документы, полностью составленные с помощью этих языков, так и предметные области, включающие элементы семантического кода. Однако в этих ИПС поиск ведется, как правило, по ключевым словам, и поэтому возникают проблемы с обнаружением онтологии, интересующей пользователя и имеющей приемлемую для него сложность.

Кроме того, основная часть ИР, представленных в заданной предметной области, не сопровождается метаданными RDF (а если и сопровождается, то доверие к этим метаданным остается открытым вопросом) или какими-либо онтологиями, а построение онтологий ИР может быть автоматизировано только частично и, в любом случае, требует участия человека на ряде этапов, оставаясь при этом достаточно длительным и трудоемким процессом.

В последние годы онтологии применяются во многих прикладных интеллектуальных информационных ресурсах. Следует отметить, что, несмотря на тот значительный интерес, который проявляют российские исследователи к самым разнообразным аспектам онтологического анализа – автоматизированному построению онтологий, их сопоставлению, пополнению и анализу, – именно вопросы доступа, хранения и создания метаописаний онтологий остаются практически не затронутыми. Впрочем, и во всем мире в этом направлении сделано недостаточно для эффективного решения задачи повторного использования онтологий.

Разработки, связанные с использованием онтологий, обеспечивают новые технологии в этой области. В заданной ПрО находится большое количество онтологий из самых разных ПрО. Но из-за сложности структуры онтологий и их большого количества пользователю тяжело не только вносить в них изменения и дополнения, но и вообще найти по тематике и уровню сложности онтологию.

Построенная таким образом онтология не только описывает предметные и проблемные области, но и задает структуры по представлению реальных объектов и связей между ними. В связи с этим данные представляются в качестве множеств взаимосвязанных ИО, которые соответствуют некоторым понятиям онтологии и имеют заданные им структуры. Семантики связей в ИПС и между ИО определяются отношениями, которые заданы между соответствующими понятиями онтологий. Совокупности таких ИО и их связей образуют информационные содержания или контент.

Тезаурус при этом обеспечивает возможности взаимодействия на нескольких языках, включая навигацию, поиск, а также обработку информационных ресурсов, которые представлены на разных языках.

Тезаурус имеет следующий вид $Th = (Tr, At, Rt, RTO, Axt)$, где $Tr = \{Tr_1, \dots, Tr_n\}$ – является конечным непустым множеством терминов, которые представляют понятие и отношение некоторой ПрО; из этого множества терминов Tr следует выделить подмножество базовых терминов $Trb \subseteq Tr$, считающиеся наиболее предпочтительными для представления имени понятия и отношения; $At = \{at_1, \dots, at_w\}$ – конечные множества атрибутов, которые описывают свойство терминов Tr ; $Rt = \{Rt_1, \dots, Rt_m\}$, $Rt_i \subseteq Tr * Tr$, $Rt = R_{SBT} \cup R_{SNT} \cup R_{AT} \cup \{R_{USE}, R_{UF}, R_{LE}, R_{TO}\}$ – конечные непустые множества бинарных отношений, которые заданы на терминах ПО согласно принятым стандартам ГОСТ и ISO [1]; R_{SBT} –

множества бинарных отношений, которые связывают некоторый термин с термином более общих понятий; R_{SNT} – множества обратных отношений к R_{SBT} ; R_{AT} – конечные множества ассоциативных отношений между терминами; R_{UF} – бинарные отношения, связывающие наиболее предпочтительные термины с синонимами на том же языке; R_{USE} – обратное отношение к R_{UF} ; R_{LE} – отношения лексических эквивалентностей между терминами, которые определяют одни и те же понятия, но на разных языках; R_{TO} – отношения, устанавливающиеся при соответствии между терминами тезауруса и понятиями или отношениями онтологии, т.е. $R_{TO} \subseteq Trb * Eo$, где Trb – множества базовых терминов тезауруса, $Eo = C \cup R$ – множественные понятия и отношения онтологий; Axt – множества аксиом, которые определяют семантику связей между всеми терминами.

Первым шагом в этом направлении являются подходы по созданию коллекций онтологий и соответствующих им ресурсов. Первоначальные проекты по сбору базы существующих онтологий предлагали создание систем управления онтологией, которые предоставляют различные функции для управления, адаптации и стандартизации групп онтологий. Эти системы являются важными средствами для группировки и реорганизации онтологий, для их дальнейшего повторного использования, интеграции, технического обслуживания, отображения и управления версиями.

Система управления онтологией представляет собой программное обеспечение для хранения, организации, изменения и извлечения знаний из онтологий, при этом поддерживает семантику для функций хранения, организации, модифицирования и извлечения знаний из заданной ПрО. Кроме того, в различных реализациях в ней могут поддерживаться и многие другие функции, связанные с обработкой онтологий. Она является аналогом системы управления базами данных (СУБД) для работы со специфичным контентом – онтологиями с учетом их семантики и основной ее целью создания является поддержка многоязычного доступа к знаниям и их повторного использования человеком и машиной.

Другой проблемой является то, что онтологии, которые пытаются представить знания достаточно широкой ПрО, оказываются слишком громоздкими для эффективного использования. Но и модульности онтологии недостаточно для повторного использования онтологий, если разработчики не могут эффективно находить нужные модули. Поэтому возникает необходимость в соответствующей инфраструктуре, которая поддерживала бы интеллектуальное исследование онтологий и их выбор конечными пользователями.

С технической точки зрения различные практические реализации онтологий значительно отличаются друг от друга. Существует проблема интероперабельности между ними, поскольку разработчики применяют разнообразные методы и технологии для интерпретации и использования метаданных. Кроме того, в большинстве существующих онтологий плохо поддерживаются такие

функции, как модуляризация и версияльность, так же, как и отношения между онтологией и средой разработки онтологий для поддержки всего жизненного цикла онтологии.

В литературе описано много альтернативных вариантов проектирования систем доступа к информационным ресурсам и поиска информации. Наиболее часто используются для этой цели оценки систем информационного поиска [1]. Например, чаще всего для оценки информационного поиска используются точность и полнота.

Пользователь может обращаться к онтологиям, созданным другими пользователями – пересматривать их, задавать по ним контекст поиска, копировать из них нужные фрагменты, но не имеет права изменять их. ИПС может обеспечить поиск онтологий, которые содержат введенные пользователем термины, а также поиск онтологий, похожих на выбранную пользователем онтологию. Это позволяет создавать группы пользователей с общими информационными интересами и предотвратить дублирование в выполнении одинаковых многоразовых запросов разных пользователей.

Одна из основных проблем связана с тем, что пользователю для описания ПрО своих информационных потребностей приходилось затрачивать достаточно значительные усилия на создание соответствующей онтологии либо описывать ее недостаточно точно, пользуясь одной из предложенных онтологий.

При использовании внешних онтологий эта проблема хотя и не решается полностью, но в значительной мере упрощается – найти достаточно близкую интересам пользователя онтологию позволяют как объем контента, так и наличие стандарта для описания семантики представленных там онтологий. Кроме того, они обеспечивают ряд качественных оценок хранящихся в них онтологий. Поэтому интегрирование онтологий позволяет значительно расширить область применения поисковой системы, обеспечить ее эффективность при распознавании различных типов информационных объектов и применять совместно с различными интеллектуальными рекомендующими системами с многоязычным доступом.

Предложенный подход обеспечивает многоязычный содержательный доступ к знаниям и информационному ресурсу заданной предметной области на основе совместных использований онтологий и тезаурусов. Существующие связи между терминами и понятиями, обеспечиваются визуализацией информации на различных языках, а также создаются предпосылки для их совместного использования во время поиска и обработки информации.

Актуальность проблемы разработки и использования онтологий и тезаурусов подтверждается рассмотренными в статье исследовательскими инициативами. В то время как онтологии являются механизмом для интероперабельности и обмена данными между информационными объектами, сами онтологии почти всегда создаются изолированно. Не сформулировано никакого

общего представления об аннотации для онтологии посредством метаданных и никаких общих способов идентификации версии онтологии. Различные онтологии используют разнообразные технологии и методы аннотирования и редактирования онтологий, не ограниченные какими-либо стандартными соглашениями, на протяжении всего их жизненного цикла. Для решения этих проблем важно достигнуть интероперабельности между онтологиями через общие интерфейсы, стандартные форматы метаданных.

В настоящее время активно проводятся научные исследования в направлении организации связей разработки стандартов для метаописаний онтологий. Решение этой проблемы позволило бы осуществлять глобальный поиск онтологий и их компонентов не только по ключевым словам, а на уровне их семантики, что должно обеспечить многоязычный доступ к повторному использованию представленных онтологий заданной ПрО. Практика интеграции существующих онтологий с семантическими поисковыми системами позволит сформулировать требования к новым стандартам онтологических метаданных и формам их представления.

Литература

1. **Гладун А. Я.** Онтологии как инструмент распознавания интеллектуальных информационных объектов в распределенных приложениях и системах / А.Я. Гладун, Ю.В. Рогушина // Труды XVIII-ой Международной научно-практической конференции «Информационные технологии в экономике, менеджменте и бизнесе». – К.: Европейский университет, 2012. – С. 51-55.
2. **Журавлев Ю. И.** Об алгебраическом подходе к решению задач распознавания и классификации / Ю.И. Журавлев // Проблемы кибернетики. Вып. 33. – М.: Наука, 1978. – С. 5-68.
3. **Величко В.** Структурирование онтологии ассоциаций для конспектирования естественно-языковых текстов / В. Величко, В. Гладун, Л. Святогор // International Book Series, N.2. Advanced Research in Artificial Intelligence. Supplement to the International Journal “Information Technologies & Knowledge”. – 2008. – V. 2. – 153 p.
4. **Ефименко И. В.** Онтологическое моделирование экономики предприятий и отраслей современной России: Часть 1. Онтологическое моделирование: подходы, модели, методы, средства, решения / И.В. Ефименко, В.Ф. Хорошевский // Математические методы анализа решений в экономике, бизнесе и политике. – М.: Изд. дом Высшей школы экономики, 2011. – 76 с.
5. **Guarino N.** Formal Ontology in Information Systems // Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press. – P. 3-15.
6. **Загорюлько Ю. А.** Построение порталов научных знаний на основе онтологий // Вычислительные технологии. – 2010. – Т 12. – Спец. вып. 2. – С. 169-177.
7. **Dagobert S.** Multilingual thesauri and ontologies in cross-language retrieval // AAAI Spring Symposium on Cross-Language Text and Speech Retrieval. – Stanford: Stanford University, 2007. – 306 p.