

УДК 004.8

РАЗРАБОТКА ОНТОЛОГИЧЕСКОЙ МОДЕЛИ ИНФОРМАЦИОННОЙ ПОТРЕБНОСТИ ПОЛЬЗОВАТЕЛЯ ПРИ СЕМАНТИЧЕСКОМ ПОИСКЕ

Ю.В.Рогущина

*Институт программных систем НАН Украины, Киев, Украина
ladamandraka2010@gmail.com*

Аннотация

В статье рассматриваются разработка онтологической модели информационной потребности пользователя, для удовлетворения которой производится семантический поиск. Предложена ее структура (основные компоненты – пользователи и их персональные программные агенты, информационные ресурсы, онтологии предметных областей, информационные объекты и т.д., а также связи между этими компонентами), средства пополнения и усовершенствования этой модели, включая метод ее использования при семантическом поиске. Обосновывается целесообразность онтологического подхода к построению такой модели, отображающей как персональную (об отдельных пользователях), так и коллаборативную (о группах пользователей) информацию. Для прогнозирования поведения поисковой системы предлагается использовать парадигму интеллектуальных программных агентов.

Ключевые слова: семантический поиск, онтологическая модель, тезаурус, индуктивный вывод, рекомендующие системы, коллаборативный поиск

Введение

Проблема эффективного удовлетворения информационных потребностей (ИП) пользователей является сегодня одной из ключевых при разработке самых разнообразных информационных систем (ИС), связанных с управлением знаниями [1]. Это обусловлено тем, что большинство современных приложений ориентированы на использование и преобразование знаний об интересующей пользователя предметной области (ПрО), которые закладываются в ИС при её разработке, а извлекаются динамически из доступных информационных ресурсов (ИР). Как правило, доступ к различным ИР обеспечивается через Web, где для их представления могут использоваться самые разные модели формализации, способы и форматы хранения, условия доступа и методы обработки.

Таким образом, проблема извлечения знаний из Web оказывается составной частью самых разнообразных ИС и в общем случае связана с распознаванием информационных объектов (ИО), имеющих отношение к решаемой пользователем задаче. Эта проблема может рассматриваться как семантическая (использующая знания) надстройка над традиционным поиском информации.

В наиболее общем виде решаемая в данном исследовании проблема заключается в следующем: чтобы эффективно удовлетворить потребность пользователя в необходимой ему информации, необходимо получить формализованное описание этой потребности, опирающееся на знания о том, что пользователю надо и в какой форме. При этом пользователь, как правило, не обладает достаточными навыками для создания формализованных описаний задачи и ПрО (например, в виде онтологии), а если и обладает, то часто не соглашается тратить много времени и усилий на создание таких описаний. Поэтому необходимо формировать такие описания автоматизировано.

Кроме того, использование слишком сложных формализмов для таких описаний, хотя и обеспечивает адекватную математическую модель для поиска, но не позволяет пользователю самостоятельно понять её и обнаружить её несоответствие стоящим перед ним задачам. Таким образом, проблема заключается в нахождении некоего компромисса между сложностью модели и её наглядностью.

Поэтому необходимо проанализировать достоинства и недостатки существующих подходов к поиску информации и на этой основе построить модель взаимодействия пользователя с открытой информационной средой Web.

Следует отметить, что в работе рассматривается *частный случай* такого поиска, при котором пользователь согласен затратить определённые усилия для того, чтобы поиск осуществлялся более качественно. Для этого у пользователя может быть одна из ряда причин, и в каждом случае информационно-поисковая система (ИПС) должна учитывать специфику ситуации. Рассмотрим наиболее характерные ситуации:

1. Пользователи имеют стабильные долговременные ИП в тех областях, где сами они являются экспертами и достаточно хорошо представляют себе как структуру ПрО, так и структуру искомых информационных объектов. Такой поиск используется, в частности, научными сотрудниками и техническими специалистами.
2. Решение, которое пользователь собирается принять по результатам поиска, является для него очень важным, а качество получаемых результатов ему сложно оценить самостоятельно, так как он сам не обладает достаточными знаниями в данной ПрО. Такая ситуация может быть связана, например, с дорогостоящими покупками, с поиском работы или с лечением, т.е. действиями, которые пользователь предпринимает крайне редко. В этом случае может быть полезен опыт других пользователей, принимавших аналогичное решение, но для того, чтобы результат оказался эффективен, пользователю потребуется достаточно подробно описать как решаемую проблему, так и собственные параметры.
3. Потребности и предпочтения пользователя сложно формализовать, хотя сами оценки тем или иным информационным объектам пользователь может дать относительно легко. Такая ситуация характерна при подборе интересной для пользователя художественной литературы, музыки, украшений и предметов искусства, в оценке которых преобладают субъективные факторы. В таком случае для эффективного поиска пользователю надо предоставить набор удовлетворяющих его ИО, оценить некоторое количество ИО, предложенных ИПС, а поисковая система попытается найти общие свойства «хороших» ИО.

Следует учитывать, что, хотя пользователь может иметь интересы в различных областях и решать различные, не связанные друг с другом задачи, его собственные свойства при потреблении информации остаются достаточно стабильными, и потому ИПС может использовать эти сведения при решении новых задач.

1 Моделирование информационной потребности пользователя

В наиболее общем виде информационный поиск можно рассматривать как сопоставление модели ИП пользователя с информационными моделями различных ИР, доступных ИПС. Чтобы более точно сформулировать теорию интеллектуального информационного поиска, необходимо строго определить основные компоненты их семантической модели, их свойства и связи между ними.

Основными элементами этой модели, которые можно выделить при анализе существующих ИПС, являются:

- пользователь;
- ИП пользователя;

- запрос пользователя – явным образом переданное пользователем сообщение о наличии ИП;
- задача, для решения которой пользователь нуждается в удовлетворении своей ИП
- группа пользователей – некоторое подмножество всей совокупности пользователей ИПС, сгруппированных по какому-либо признаку;
- ИР – документы различных типов, к которым ИПС имеет доступ;
- ИО – сведения об объекте с определённой структурой, указанной пользователем, которые содержатся в одном или нескольких ИР;
- информационная среда – совокупность всех доступных ИР, их свойств (включая их оценки пользователями) и связей между ними;
- ПрО – некоторая часть ИР, связываемых в одно множество в соответствии с каким-то признаком их контента (принято считать, что понятие ПрО – множества всех предметов, свойства которых и отношения между которыми рассматриваются в научной теории – в общем случае не может быть формализовано как первичное понятие);

В дополнение к этим элементам предлагается ввести в модель еще один элемент – *агент пользователя* (АП). АП представляет собой интеллектуальный программный агент, который представляет интересы пользователя во взаимодействии с ИПС. Для описания поведения такого агента используются интенциональные отношения, с помощью которых можно формализовать цели, намерения и желания пользователя. Использование такого формализма, как АП, позволяет, с одной стороны, избежать приписывания человеку-пользователю искусственно суженной и формально описанной сферы интересов, а с другой – обеспечит средства и методы прогнозирования его поступков в рамках описываемой модели взаимодействия пользователя и ресурсов в открытой информационной среде. Кроме элементов, в модели должны отображаться связи между ИР и ПрО; ИП и ПрО; ПрО и задачами пользователей; пользователями и ПрО; пользователями из группы.

Для того чтобы формализовать описание элементов модели и обеспечить его автоматизированную обработку, целесообразно использовать онтологическое представление знаний о них. Таким образом, для описания ИП пользователя надо построить онтологическую модель, классы которой соответствуют перечисленным выше элементам, а отношения – связям между ними. Чтобы определить, какими атрибутами должны обладать классы, необходимо проанализировать историю развития ИПС и выявить тенденции их совершенствования.

1.1 Традиционный информационный поиск

При *традиционном информационном поиске* ИП представляется пользователем в виде запроса, который состоит из последовательности ключевых слов (иногда утверждают, что он представляет собой множество ключевых слов, но порядок введения слов влияет на результаты запроса, а в некоторых случаях в процессе поиска учитывается и непосредственное совпадение последовательности ключевых слов с фрагментом текстового ИР).

Получив такой запрос, ИПС сопоставляет его со сведениями о контенте проиндексированных ИР (как правило, каждый ИР – это отдельный документ), и в ответ на запрос пользователя формируется группа ИР, по тем или иным параметрам признанных ИПС соответствующими запросу, т.е.

$$I = \{i_j, j = \overline{1, n}\} = f(z, DB_{unc}).$$

Такой поиск обычно основывается на обнаружении лексического соответствия ключевых слов и терминов, содержащихся в документе, на основе минимальных предварительных знаний, рассматривая ИР как набор слов – последовательностей букв и цифр, отделенных друг от друга пробелами и другими символами-разделителями. В основе сопоставления лежит

векторная модель документов и запросов. При этом субъектами, обрабатываемыми ИПС при каждом поступившем запросе, являются пары (запрос, ИР), а их сопоставление производится унифицировано для всех запросов и всех пользователей.

Как правило, такой поиск применяют для обработки неструктурированной или слабо-структурированной текстовой информации – например, для естественно-языковых документов, в которых структура присутствует неявно и может быть выявлена лишь в результате лингвистического анализа. Кроме того, при сопоставлении часто учитывается расстояние между обнаруженными терминами – чем ближе в тексте они находятся друг к другу, тем выше степень соответствия. Для структурированных документов имеет также большое значение месторасположение ключевых слов (например, в заголовке или в метаописании).

Анализ исследований в области информационного поиска показал, что дальнейшее усовершенствование средств описания и алгоритмов сопоставления запросов и ИР практически не улучшает ситуацию, и поэтому для более эффективного поиска нужно использовать дополнительные знания – о ПрО, интересующей пользователя, о сообществах пользователей с подобными интересами и о качестве ИР.

1.2 Персонализированный информационный поиск

При *персонализированном поиске* субъектами обработки для ИПС являются тройки: запрос, пользователь, ИР. Так как при сопоставлении учитываются свойства конкретного пользователя, то результаты поиска по одному и тому же запросу оказываются различными для различных пользователей – в зависимости от их персональных способностей, интересов и предпочтений. Сведения о пользователе хранятся в БД поисковой системы и могут содержать его формальные характеристики (например, перечень естественных языков, которые знакомы пользователю), историю ранее выполненных запросов и знания об интересующих его ПрО.

$$I_{pers} = f(z, u, DB_{unc}).$$

В различных ИПС сохраняются различные сведения о пользователях. В общем случае такие сведения можно разделить на три категории:

- полученные от пользователя явно (при регистрации или в процессе опроса);
- импортированные из внешних источников информации – социальных сетей, домашних страниц, предложенных пользователем ИР, Википедии и т.д.;
- извлеченные из данных, накопленных ИПС в процессе взаимодействия с пользователем.

К сожалению, опыта конкретного пользователя не всегда достаточно для того, чтобы определить полезность тех или иных ИР (например, при обращении к новой ПрО или к новой группе ресурсов). В таком случае полезным может оказаться опыт других пользователей, производивших ранее поиск в той же области и имеющих сходные ИП. При коллаборативном поиске целесообразно использовать методы, применяемые в рекомендующих системах.

Коллаборативный поиск – поиск, при котором используются знания и опыт, полученные при совместной работе пользователей с ИПС. При этом предполагается, что пользователи объединяются в группу при наличии похожих интересов или в случае, если они решают похожие задачи.

1.3 Семантический информационный поиск

Семантический поиск предполагает, что информация в процессе сопоставления должна обрабатываться на семантическом уровне, с использованием знаний (о пользователе, ресурсах, ПрО и т.д.). Даже при самом простом варианте знание-ориентированного поиска при об-

работке запроса учитываются не только формальные сведения о запросе, пользователе и ИР, но и более сложно структурированные знания о них. Тогда при их сопоставлении необходимо будет оценивать степень подобия этих знаний. В частности, при онтологическом подходе к представлению знаний для каждого ИР может указываться онтология ПрО, характеризующая его контент, а для пользователя – онтология интересующей его ПрО, а при сопоставлении ИР и запроса необходимо будет выполнить сопоставление этих двух онтологий.

Кроме того, наличие дополнительных знаний о том, что именно ищет пользователь – например, о структуре искомого ИО – позволяет структурировать найденную информацию и предоставлять ее пользователю в более удобном виде. Например, если пользователю нужно найти информацию о каком-то учёном, и пользователь указывает, какие именно свойства этого человека представляют для него интерес, то при семантическом поиске ИПС может явным образом указать фамилию человека, его специальность, наличие публикаций и т.п.

Семантический поиск определяется как метод информационного поиска, в котором релевантность документа запросу определяется семантически (по близости смысла), а не синтаксически (по встречаемости ключевых слов в документе). От того какие именно знания используются, как они представлены и как они обрабатываются, зависит специфика разрабатываемой ИПС и её концепция, но в общем случае

$$I_s = \{i_j, j = \overline{1, n}\} = f(z, DB_{unc}, KB_{unc}).$$

При этом часто предоставляются возможности для нечёткого поиска (например, обрабатываются неправильно написанные ключевые слова) и для поиска с учётом контекста.

1.4 Специфика информационного поиска в Web

Если речь идет о семантическом поиске в Web, то следует учитывать, что при этом в Web могут находиться не только ИО, среди которых осуществляется поиск, но и внешние базы знаний (БЗ), используемые при поиске. Поэтому при создании таких систем следует учитывать, что такие внешние БЗ могут менять контент, структуру и доступность независимо от разработчиков ИПС.

$$I_{web_s} = \{i_j, j = \overline{1, n}\} = f(z, DB_{unc}, KB_{unc}, \{KB_{web_k}, k = \overline{1, m}\}).$$

Следует учитывать, что сегодня многие ИПС (например, Google) стремятся накапливать и использовать опыт взаимодействия с конкретным пользователем. Но часто ИП пользователя оказываются ограниченными во времени (например, накопив информацию для выбора нового телефона, пользователь покупает его и больше не нуждается в сведениях о телефонах, а ИПС продолжает предлагать их ему) либо вообще не связанными с ним (например, запрос выполняется по чьей-либо просьбе). Кроме того, часть своих ИП пользователь не хочет делать открытой информацией, например, запросы, связанные с отдыхом или здоровьем не хочет смешивать с запросами по работе. Поэтому более целесообразно при выполнении запросов дать возможность пользователю включить его в один из своих профилей, либо вообще не сохранять для дальнейшей обработки.

1.5 Тенденции развития информационного поиска

На основе анализа существующих ИПС можно выделить основные тенденции развития поиска [2]:

- от формального – к семантическому;
- от унифицированного – к персонифицированному;
- от индивидуального – к коллаборативному;
- от закрытого – к управляемому;

- от монотонного – к тематическому (с учётом динамики и конечности ИП).

Наиболее полно удовлетворить ИП пользователя позволяет интегрированное использование всех этих возможностей, т.е. персонифицированный, Web-ориентированный коллаборативный поиск, основанный на знаниях. Как правило, такой поиск является надстройкой над уже существующими поисковыми системами (например, Google), которая позволяет перепорядочить результаты поиска.

2 Постановка задачи

Чтобы обеспечить для различных интеллектуальных приложений эффективный доступ к ресурсам открытой информационной среды Web, необходимо разработать интегрированную формальную модель, обеспечивающую интероперабельное представление знаний о пользователях, ресурсах и специфике ПрО (в частности, учитывающих Semantic Web и социальный Web). При разработке методов обработки, представленных в этой модели знаний, необходимо проанализировать уже существующие средства онтологического анализа, возможности использования тезаурусов, методы индуктивного извлечения знаний и алгоритмы выработки рекомендаций на основе накопленного сообществом пользователей опыта об интересующей пользователя ПрО.

3 Использование онтологического анализа для семантического поиска

Различные действия, связанные с извлечением и обработкой знаний, представляют собой крайне трудоемкие, сложные и полностью не автоматизируемые процессы. Вследствие этого знания являются достаточно дорогостоящим продуктом. Поэтому возникает потребность в обеспечении повторного использования уже извлеченных знаний и в их интероперабельном представлении для этого. Основой для этого может послужить *онтологический анализ*, базирующийся на стандартах, технологиях и программных средствах, разработанных в рамках проекта Semantic Web [3].

То, что онтологии являются адекватным средством для описания различных ПрО, является на сегодня общепризнанным фактом, а широкий выбор онтологий, доступных через Web, подтверждает популярность этого подхода среди различных групп разработчиков и пользователей Web-приложений.

Кроме того, ряд языков описания онтологий (например, OWL DL) базируется на дескриптивных логиках, что обеспечивает достаточно точное прогнозирование сходимости и времени работы различных методов обработки знаний, представленных в виде онтологии.

Для того, чтобы повысить эффективность поиска, целесообразно как можно более точно определить два свойства ИП пользователя – 1) к какой ПрО относится ИП пользователя и 2) какой именно тип информации необходим для её удовлетворения. В первом случае пользователю нужно выбрать одну из существующих онтологий ПрО и при необходимости модифицировать её в соответствии со спецификой его проблемы. Так как не только создание и модификация онтологий, но и анализ содержащихся в уже созданных онтологиях знаний представляет собой достаточно сложную задачу, то такой подход приемлем только в том случае, если в дальнейшем такая онтология будет использоваться для выполнения не единичного запроса, а для достаточно большого числа связанных с этой ПрО запросов.

В последнее время многие разработчики ИПС в той или иной степени декларируют применение онтологий и поддержку семантического поиска. Например, семантический поиск Google включает три основных компонента [4]: адрес URI; RDF; значение (онтологию). При этом онтологии позволяют связывать друг с другом различные данные, описанные при по-

мощи RDF и адресованные через URI. Кроме того, для поддержки семантического поиска в Google применяется новый инструмент, обеспечивающий семантический поиск, – Knowledge Graph (Сеть знаний), который обеспечивает связь между различными элементами проиндексированного контента ИР и позволяет объединять информацию из различных источников.

Knowledge Graph позволяет непосредственно на странице с результатами поиска получать информацию об объекте поиска и связанные с ним факты: справа от результатов поиска на экран выводится информационная панель, отображающая сведения о географических объектах, людях, фильмах и т.п. Это позволяет пользователю получить информацию, не переходя на сайт, послуживший источником информации.

Knowledge Graph не просто идентифицирует ключевые слова, но и анализирует семантику поискового запроса и ищет информацию в базе данных Google, которая содержит сведения о различных объектах – людях, местах, предметах и т.п.

Google использует правила вывода, чтобы группировать информацию на основе того, как пользователи работают с данными в социальных сетях (распространяемый контент, комментарии и взаимодействие), форумах и при поиске в Google. Кроме того, отслеживается поведение пользователей на различных Web-сайтах и анализируются последовательности их действий. Важным источником информации служат и сервисы, определяющие географическое местонахождение пользователей по сигналу GPS от их мобильных устройств и по IP-адресу.

Разработчики Google утверждают, что персонализация поиска в сочетании с семантическим поиском повышают релевантность результатов поиска. Тем не менее, следует отметить, что сейчас такая информация для запросов – не только на русском или украинском языках, но и на английском – предоставляется в ответ только на самые короткие и простые запросы и по охвату значительно меньше, чем набор статей в Википедии. Поэтому остается открытым вопрос, насколько будет полезен пользователям новый сервис Google и насколько на самом деле в нём задействован семантический поиск.

Главная проблема, с которой сталкиваются пользователи Knowledge Graph, – отсутствие явной модели поиска и сведений о структуре и заполнении онтологии, используемой при поиске, которая позволяла бы прогнозировать его результаты. Кроме того, непонятно, какую политику применяет Google в тех случаях, когда ответы не однозначны и зависят не только от местоположения пользователя, языка, на котором он говорит, времени года и т.п., но и от более сложных и субъективных факторов (например, запрос «самые красивые собаки» или «выдающиеся политики», на которые не может быть однозначного ответа).

4 Онтологическая модель взаимодействия пользователей и ресурсов в Web

Так как семантический поиск – это надстройка над традиционным информационным поиском, в которой для повышения pertinентности поиска используется обработка знаний как о самом пользователе и его ИП (персонализация поиска), так и об ИР, среди которых осуществляется поисковая процедура, то для его осуществления необходимо сформировать набор формальных моделей его компонентов и указать связи между ними: информационных моделей пользователя; интересующей его ПрО; задачи, которую решает пользователь; информационных моделей доступных ИР, которые характеризуют их семантику, и их дальнейшего сопоставления (рисунок 1).

Сейчас для интероперабельного представления различных знаний в Web широко применяются онтологии, обеспечивающие явное формализованное представление семантики представленной информации и обеспечивающие возможность логического вывода на них. Поэтому представляется целесообразным для поддержки семантического поиска разрабатывать

именно онтологические модели взаимодействия пользователей и ИП в информационном пространстве Web, а также методы их сопоставления и пополнения.

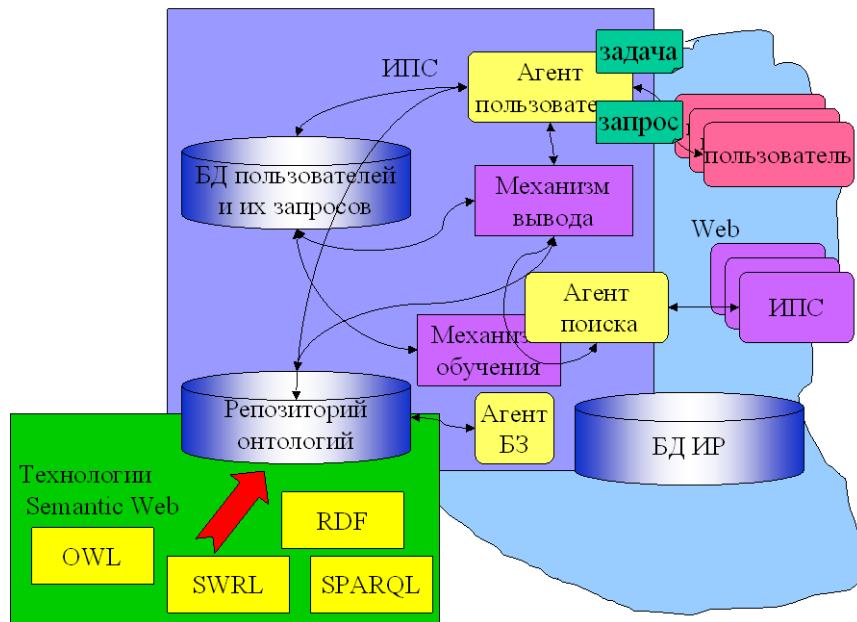


Рисунок 1 - Использование технологий Semantic Web в семантическом поиске

Онтологическая модель пользователя представляет собой класс онтологии, экземплярами которого являются сведения о зарегистрированных в ИПС пользователях (рисунок 2). В ней описаны следующие классы:

– *онтология* PrO , которая описывает область, к которой относятся ИП пользователя $O_{PrO_i} = \langle T_{PrO_i}, R_{PrO_i}, F_{PrO_i} \rangle, i = \overline{1, n}$;

– *лексическая онтология* PrO , которая содержит сведения о лексемах естественных языков, соответствующих терминам онтологии PrO $O_{lex_i} = \langle T_{lex_i}, R_{lex_i}, f \rangle, i = \overline{1, n}$, т.е. $\forall t \in T_{PrO_i}$;

– *тезаурус задачи* – множество пар, первым элементом в которых являются термины онтологии, совокупность которых характеризует ту конкретную задачу из PrO , которую в данный момент решает пользователь, а вторым – вес (положительный или отрицательный) этого термина для данной задачи $Th_{i_j} = \langle \{th_{k_{ij}} \in T_{PrO_i}, v_{k_{ij}}\}, k = \overline{1, s_{ij}}, j = \overline{1, m_i} \rangle$;

– *запрос* – множество ключевых слов, характеризующих одну из ИП пользователя, связанный с конкретной задачей при помощи тезауруса $z = \langle \{k_q\}, Th_{i_j} \rangle, q = \overline{1, u}$;

– *тема* – множество запросов, связанных с одной ИП $thema = \langle id_{thema}, \{z_q\} \rangle, q = \overline{1, u}$, которое может объединять запросы разных пользователей, базирующиеся на различных онтологиях и тезаурусах, и позволяющее объединять семантически связанные запросы;

– *результат запроса* – множество пар, первым элементом в которых являются ссылки на ИП, а вторым – оценки этих ИП пользователем $rez = f(z, u) = \langle \{id_{ir}, rating_{ir}\} \rangle$;

– *пользователь* – класс, имеющий более сложную структуру и имеющий следующие атрибуты, которые можно разделить на несколько групп.

1. Регистрационная информация:

- идентификатор пользователя;
- пароль для доступа к ИПС.

2. Опыт взаимодействия ИПС с пользователем:

- список онтологий, которые пользователь применял для описания своих информационных интересов;
- список тезаурусов, которые пользователь применял в поисковых запросах;
- список ранее выполненных запросов;
- список результатов выполненных запросов с оценками пользователя для найденных результатов.

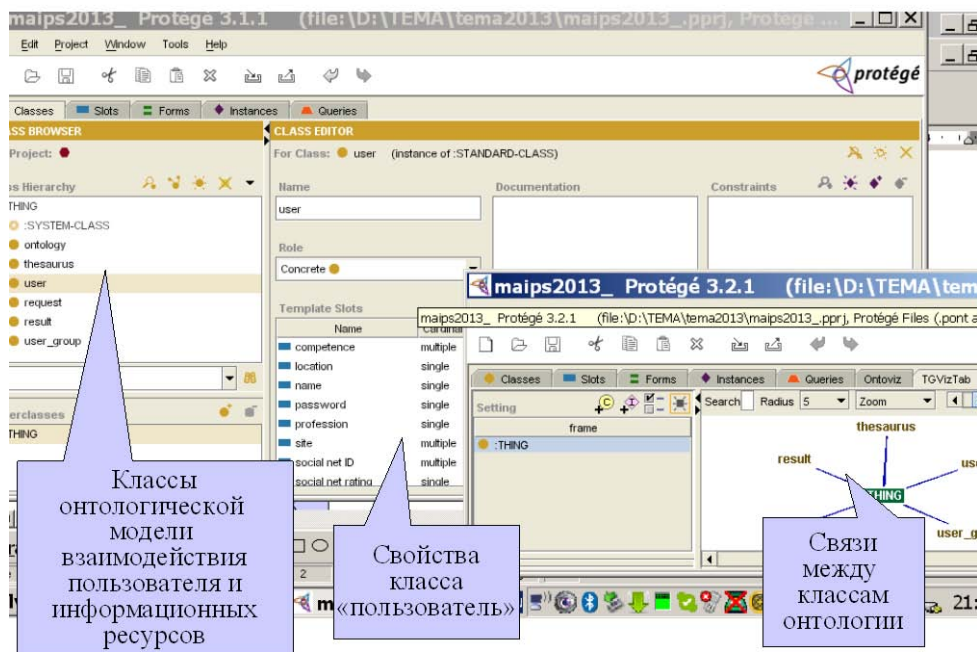


Рисунок 2 - Онтологическая модель взаимодействия пользователей и ИР

3. Сведения, импортируемые из внешних источников (необязательные сведения, могут отсутствовать):

- идентификаторы пользователя в социальных сетях, позволяющие динамически обновлять сведения о нём; рейтинги пользователя в социальных сетях;
- адрес пользователя в Википедии и других вики-ресурсах;
- адрес сайта пользователя;
- сфера компетенций пользователя (ключевые слова, импортируемые из социальных сетей);
- ссылки на публикации пользователя.

4. Собственные характеристики пользователя:

- сфера компетенций пользователя (список ключевых слов, вводимых пользователем непосредственно);

5. Формальные данные о пользователе (необязательные сведения, позволяющие ИПС формировать группы пользователей со схожими ИП):

- место жительства;
- возраст;
- профессия, образование и т.д..

– группа пользователей – класс, свойствами которого являются идентификатор группы и список пользователей, по тем или иным причинам объединённых в одну группу (группы могут формироваться явно путём выбора пользователя или автоматически на основе соответ-

вия каким-либо условиям, например, группы пользователей со сходными формальными данными или выполняющих похожие запросы) $gr = \langle id_{gr}, \{u_i\}, i = \overline{1, n} \rangle$;

– *информационный ресурс* – сведения о найденных ранее ресурсах и их оценках $\langle U_{url}, \{z_i, m_i, q_i\}, i = \overline{1, n} \rangle$, включающие идентификатор ресурса, запросы, по которым он был обнаружен, оценку пользователя, которому он был предоставлен, и его уровень читабельности для этого пользователя.

5 Использование в семантическом поиске методов рекомендуемых систем

Как известно, *рекомендующие системы* (РС) отличаются от ИПС тем, что пользователю не надо явным образом формулировать поисковый запрос – система сама, на основании имеющихся сведений о пользователе, предлагает ему рекомендуемые элементы (РЭ) [5]. Персонализированные рекомендации, вырабатываемые такими системами, – это упорядоченные списки РЭ, т.е. работа РС сводится именно к ранжированию доступных РЭ. Чем выше и точнее информированность РС о потребностях пользователя, тем более эффективны результаты её работы. Но даже в том случае, если пользователь явно формулирует свои ИП, с помощью методов, используемых в РС, можно значительно повысить пертинентность поиска, персонифицируя запрос и используя коллаборативные знания о различных сообществах пользователей, сгруппированных на основе подобных ИП [6].

Формально создание рекомендаций в РС может быть представлено следующим образом. Пусть C – множество пользователей РС, S – множество предлагаемых РЭ (товаров, книг, фильмов, сервисов и т. д.), U – функция полезности, описывающая интерес пользователя $c \in C$ к РЭ $s \in S$, т. е. $U : C \times S \rightarrow R$, где R – количественная оценка. Цель РС – для каждого потребителя $c \in C$ выбрать такой РЭ $s' \in S$, что $U(c, s') = \max_{s \in S} U(c, s)$. Каким именно образом определяется функция полезности, зависит от типа РС и от специфики РЭ.

В РС при выборе РЭ используют:

- *персональный* подход – анализ профиля конкретного пользователя, его ранее проявленных предпочтений и явным образом выраженных условий;
- *коллаборативный* подход – анализ предпочтений других пользователей, которые по тем или иным причинам могут распространяться и на того пользователя, для которого делается выбор;
- *контент-ориентированный* подход, при котором анализируются сами РЭ, предлагаемые пользователю;
- *доверительный* подход – анализируется качество предлагаемых пользователю РЭ и анализируется степень доверия к ним.

Следует отметить, что в большинстве реальных РС все эти подходы реализуются интегрировано, но им придаётся различное внимание.

В РС подобие между двумя пользователями основывается на том, какие оценки они дали одним и тем же РЭ. Для эффективной работы РС надо предвидеть оценки, исходя из небольшого количества примеров. Для преодоления проблемы разреженности оценок следует при поиске похожих пользователей использовать также сведения из их профилей и обнаруживать пользователей со схожими профилями, например, относящихся к одному демографическому сегменту.

Анализ основных направлений развития современных РС [7] связывает их с использованием онтологий для представления знаний как о пользователях, так и о РЭ. При персональном подходе РС необходимо накопить достаточно сведений о пользователе, чтобы в даль-

нейшем их обобщать и анализировать. Рекомендация относительно новых РЭ для пользователей может формироваться на основе его сравнения с подобными РЭ (фильтрация на основе контента), отзывов об РЭ в сообществе пользователей (коллаборативной фильтрации), семантических отношений между РЭ (эвристические рекомендации) или сочетания этих подходов.

Коллаборативная фильтрация использует рейтинги, предоставляемые сообществом пользователей, чтобы рекомендовать РЭ конкретному пользователю. Существуют два взаимодополняющих подхода к коллаборативной фильтрации: на основе пользователя или на основе РЭ. При коллаборативной фильтрации на основе пользователя находят группы подобных пользователей, а затем конкретному пользователю рекомендуют те РЭ, которые понравились другим пользователям из той же группы.

На основе вышеприведенного анализа можно предложить следующие подходы к работе РС с использованием онтологий: формирование модели пользователя; формирование модели РЭ; создание онтологии РЭ; накопление сведений об экземплярах РЭ и экземплярах пользователей; накопление оценок РЭ пользователями; анализ экземпляров РЭ; классификация (или кластеризация) пользователей на группы с подобными интересами; формирование набора стратегий, которые пользователь может явно выбирать для получения рекомендации; построение метода, позволяющего уточнить класс необходимого пользователю РЭ.

Рассмотрим также, оценки каких именно групп пользователей целесообразно применять для коллаборативной фильтрации. Самый простой случай группы – это группа, состоящая всего из одного пользователя, для которого и осуществляется поиск рекомендаций. Можно сказать, что при этом коллаборативная фильтрация сводится к персональной. Противоположный случай – когда для пользователя значимы оценки РЭ всем сообществом в целом. Это может иметь место для тех ПрО, к которым пользователь обращается впервые и ещё не имеет собственного мнения не только о самой области, но и о критериях нахождения в ней экспертов. К промежуточным случаям относится анализ оценок экспертов в ПрО (целесообразно предоставить пользователю возможность явно задавать приемлемый уровень их квалификации).

В целом следует оценивать выбранную стратегию рекомендации по трём направлениям – учёт мнения самого пользователя, учёт мнения сообщества, анализ самого РЭ. В таком трёхмерном пространстве можно разместить большинство типичных объектов рекомендации. Выбор пользователем значения по каждому из трёх параметров для стратегии рекомендации для интересующих его РЭ и наличие онтологии (или хотя бы таксономии РЭ) позволяет достаточно точно профилировать интересы самого пользователя, оценить его собственную компетентность для оценивания РЭ (и, соответственно, значимость его мнения для других пользователей) и выявлять группы пользователей со сходными интересами. Следует отметить, что более объективным является признание своей некомпетентности в оценивании РЭ, чем декларирование своей высокой квалификации.

6 Тезаурусы как средство представления онтологических знаний

Онтология ПрО имеет достаточно сложную структуру, но при удовлетворении ИП пользователя для решения конкретной задачи ИПС использует только часть содержащихся в такой онтологии знаний. Поэтому целесообразно использовать для моделирования знаний пользователя об интересующей его ПрО *тезаурус* – частный случай онтологии, который можно рассматривать как проекцию онтологии на задачу [8, 9].

Тезаурус – это $Ts = \langle T, R \rangle$, где T – множество терминов, а R – множество отношений между этими терминами. Множества T и R конечны. Множество терминов тезауруса T со-

ответствует множеству концептов X онтологии O . Тезаурусы позволяют моделировать знания как о пользователях, так и о тех ресурсах, которые они ищут [10].

Чтобы формализовать область своих интересов – ПрО поиска – пользователю надо создать тезаурус, моделирующий интересующую его ПрО, в котором содержатся основные термины ПрО и связи между ними. Тезаурус можно создать вручную или автоматизировано. Основой для автоматического создания тезауруса может послужить обработка набора ИР, релевантных этой ПрО, или ранее созданная онтология ПрО, из которой пользователь отбирает только необходимые ему термины. Все эти подходы могут комбинироваться друг с другом.

Для создания тезаурусов ИР и РЭ предлагается использовать упрощенный алгоритм построения тезауруса: по полному перечню слов, используемых в ИР, строится словарь терминов, из которого отбрасываются стоп-слова, содержащиеся в специально разработанном пользователем списке. Этот алгоритм применяется только для тех ИР, которые не сопровождаются метаописаниями. В противном случае из метаописаний (в формате RDF или OWL) извлекаются термины тезауруса и связи между ними, которые дополняют построенный по контенту ИР словарь. Аналогично строятся тезаурусы РЭ – обрабатываются их метаописания, контент, отзывы о них других пользователей.

Пользователь вводит запрос, приблизительно идентифицируя свою ИП с помощью ключевых слов или выбирая класс интересующего его РЭ (возможно, с набором условий и ограничений), например, РЭ класса «художественная литература/фантастика/фэнтези», изданная после 2005 года. В ответ РС формирует набор РЭ, доступных системе и соответствующих этому приблизительно запросу n ссылок на РЭ и их кратких описаний $I = \{Ref_j, D_j\}$, $j = \overline{1, n}$. Здесь Ref_j – ссылка на соответствующий РЭ IP_j (или его описание), а d_j – информация об этом РЭ, доступная РС.

Если множество I не пусто, причем РС найден в ответ на запрос более чем один РЭ ($n \geq 1$), то нужно установить порядок, в каком предлагать пользователю сведения о найденных РЭ. Тогда для всех РЭ из этого множества $I = \{Ref_j, D_j\}$, $j = \overline{1, n}$ формируются их упрощенные тезаурусы $Ts(IP_j) = \langle T_j, \emptyset \rangle$, $j = \overline{1, n}$ и соответствующие им словари терминов $T_j = \{t_{j_w}\}$, $j = \overline{1, n}$, $w = \overline{1, q_j}$. t_{j_w} – это слова, которые используются в информации о j -м РЭ, найденном РС, т. е. в D_j , $j = \overline{1, n}$. q_j , $j = \overline{1, n}$ – это количество различных слов, используемых в описании D_j , $j = \overline{1, n}$. Если слова в описании повторяются, то в словаре терминов они фиксируются только один раз.

Затем пользователь формирует тезаурус интересующей его ПрО (или указывает на ранее сформированный тезаурус) $Ts_{ПрО}$ и соответствующий ему словарь терминов этой ПрО $T_{ПрО} = \{t_m\}$, $m = \overline{1, q}$. $T_{ПрО}$ – это множество, состоящее из m терминов, относящихся к интересующей пользователя ПрО. Это множество строится аналогично словарю терминов РЭ и обычно формируется как объединение словарей терминов, содержащихся в документах, которые пользователь нашел ранее и посчитал релевантными интересующей его ПрО (как в их контенте, так и в метаописаниях).

Производится сравнение $T_{ПрО}$ и T_j , $j = \overline{1, n}$, высчитывается коэффициент их близости

$$K_j = \sum_{m=1}^q \sum_{w=1}^{w_j} f(t_{j_w}, t_m), \quad m = \overline{1, q}, \quad w = \overline{1, w_j}, \quad \text{где } f(t_1, t_2) = \begin{cases} 0, & \text{если } t_1 \neq t_2, \\ 1, & \text{если } t_1 = t_2. \end{cases}$$

с собой количество терминов, встретившихся как в тезаурусе РЭ, так и в тезаурусе ПрО.

Найденные ИР упорядочиваются в зависимости от значений K_j , и пользователю предъявляются в первую очередь те ИР, которые имеют наиболее высокий коэффициент близости к ПрО.

Представляется целесообразным использовать онтологию ПрО и выделять группы слов тезауруса, соответствующих одному термину. Для этого пользователь должен связать РЭ словаря терминов тезауруса ПрО с одним из терминов онтологии ПрО, т. е. $\forall t_m \in T_{\text{ПрО}}, m = \overline{1, q}$ задать функцию $g(t_m) \in X$, которая используется для вычисления коэффициента близости K^O :

$$K^O_j = \sum_{m,l} f(t_{j_w}, t_m), \quad m = \overline{1, q}, \quad w = \overline{1, w_j}, \quad \text{где } f(t_1, t_2) = \begin{cases} 0, & \text{если } g(t_1) \neq g(t_2), \\ 1, & \text{если } g(t_1) = g(t_2). \end{cases}$$

K^O представляет собой количество терминов, которые встретились как в тезаурусе РЭ, так и в тезаурусе ПрО и при этом ссылаются на один и тот же термин онтологии ПрО. Он позволяет использовать меньший объем документов для построения тезауруса ПрО, но требует больше время для вычислений.

При создании тезауруса ПрО, которая интересует пользователя РС, необходимо явно указать основные понятия ПрО и связи между ними. К сожалению, большинству пользователей достаточно сложно это сделать (даже имея соответствующие знания и применяя их в своей деятельности). На первом этапе формирования тезауруса пользователь может выбрать одно из следующих решений: 1) самостоятельно построить с помощью одного из редакторов онтологий онтологическое описание интересующей его ПрО; 2) найти (например, в Web) онтологию на языке OWL, которую описывает ПрО, близкую к области его информационных интересов; 3) сформировать множество понятий ПрО, содержащее наиболее характерные слова и словосочетания, встречающиеся в интересующих его ИР.

7 Использование индуктивного вывода для пополнения онтологий

При построении онтологий и тезаурусов ПрО, а также при коллаборативном подходе к поиску важно определить, какие связи между терминами ПрО являются существенными для описания ИП пользователя. Пользователю достаточно сложно самостоятельно обнаружить все важные закономерности и отбросить несущественные.

Для их выявления можно воспользоваться методами индуктивного и традуктивного извлечения знаний из данных. Существуют независимые подходы к реализации подобных методов: ID3, ACLS, CART и т. д. Наиболее интересным, в связи со спецификой проводимой работы, оказался алгоритм ID3 [11], который специально разработан для извлечения ценной информации из больших объёмов слабо структурированных данных.

Невозрастающий алгоритм ID3 предназначен для обобщения опыта экспериментов, параметры и результаты которых описаны через качественные оценки (лингвистические переменные). Он обеспечивает построение бинарного дерева решений, а это недостаточно удобно для представления закономерностей многих ПрО. Поэтому для пополнения онтологий предлагается использовать ID3m [12] – модификацию ID3 для произвольного (конечного) количества решений. Он также принадлежит к невозрастающим алгоритмам. В данном случае, примерами обучающей выборки являются РЭ, доступные РС, а параметрами, по которым

они описываются, являются их свойства, описанные в метаданных и в онтологии РЭ, значения параметров оценивших их пользователей, а также термины тезауруса пользователя.

8 Семантический поиск в МАИПС

МАИПС – это мультиагентная ИПС, которая обеспечивает семантический поиск различных ИО на основе знаний, содержащихся в онтологиях. Она предназначена для поиска информации в описанных пользователем относительно узких ПрО, связанных с профессиональными или научными интересами пользователей, и рекомендует пользователю те результаты поиска, которые относятся к интересующей его ПрО и соответствуют его персональным ИП. Её можно рассматривать как рекомендующую систему, ориентированную на формирование коллаборативных рекомендаций относительно естественно-языковых и мультимедийных информационных ресурсов, доступных через Web.

Система МАИПС ориентирована на пользователей, имеющих в сети постоянные информационные интересы и требующих постоянного поступления соответствующей информации. Для этого МАИПС позволяет сохранять и повторно выполнять запросы, учитывая реакцию пользователя на ранее предложенные ему ИП (персональная фильтрация), отслеживать появление аналогичных запросов у других пользователей (коллаборативная фильтрация), сохранять формальное описание области интересов пользователя в виде онтологии (семантическая фильтрация) и т.д.

Кроме того, в МАИПС при профилировании пользователей используется специфичный для естественно-языковых ИП критерий оценивания – сложность текста для понимания. Особенностью системы является использование оригинального знание-ориентированного алгоритма, позволяющего определить сложность понимания текста для конкретного пользователя (для этого используются тезаурусы ПрО, интересующих пользователей) [13].

МАИПС базируется на технологиях Semantic Web, в частности, использует язык представления онтологий OWL и средства его обработки. Для представления знаний об интересующей пользователя ПрО используются онтологии – как созданные самими пользователями, так и найденные в репозиториях онтологий, доступных через Web [14], – и тезаурусы ПрО. При этом тезаурус строится пользователем по соответствующей онтологии самостоятельно, а онтология выбирается из набора предложенных на сайте.

Пользователь МАИПС может обращаться к онтологиям, созданными 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 многообразных запросов разных пользователей.

Онтологическая модель, описывающая семантику взаимодействия пользователей и ресурсов МАИПС в информационном пространстве Web, обеспечивает знания для выполнения следующих действий, связанных с поиском информации и основанных на рассмотренных выше методах.

1. *Предварительный этап*, когда в систему вводятся сведения об окружающем мире, т.е.:

- создается онтологическая модель, описывающая структуру информации относительно основных элементов, с которыми работает система (пользователей, ресурсов, результатов поиска и т.д.);

- вводятся онтологии ПрО, которые могут быть полезны для поиска, и ссылки на внешние репозитории и средства поиска онтологий, которые пользователь может применить для работы в специфических ПрО;
- по имеющимся онтологиям создается несколько примеров тезаурусов, которые могут быть использованы при поиске.

2. *Этап регистрации пользователя*, на котором пользователь вводит сведения, необходимые для создания нового экземпляра класса «пользователь».

3. *Этап создания нового запроса* пользователя, на котором пользователю последовательно нужно выполнить следующие действия:

- выбрать базовую онтологию, знания которой обеспечат семантическую обработку запроса;
- по выбранной онтологии создать тезаурус запроса одним из следующих способов:
 - * выбрать несколько классов или экземпляров классов из базовой онтологии;
 - * выбрать несколько классов из базовой онтологии и классы, находящиеся от них на семантическом расстоянии не более указанной пользователем величины;
 - * выбрать несколько классов из базовой онтологии и их надклассы и подклассы указанной пользователем глубины;
 - * выбрать несколько классов из базовой онтологии и классы, связанные с выбранными классами выбранным пользователем отношением, специфичным для ПрО;
 - * вручную ввести термины тезауруса, характеризующие задачу пользователя;
 - * над построенными ранее тезаурусами применить теоретико-множественные операции объединения, пересечения и дополнения;
 - * указать вес каждого из терминов тезауруса, который отражает его важность для конкретной задачи.

- создать список ключевых слов, характеризующих конкретный информационный запрос, и соединить его с одним из ранее созданных тезаурусов;
- если нужно, присоединить запрос к одной из ранее построенных групп запросов или создать для него новую группу.

4. *Выполнение запроса*, в процессе которого поисковый запрос по ключевым словам перенаправляется внешней ИПС (Google), затем МАИПС получает найденные результаты и переупорядочивает их в соответствии с количеством найденных в них терминов тезауруса и их весом. Кроме того, для упорядочения могут учитываться другие свойства ИР, например, если пользователь указывает желаемый уровень читабельности, то этот параметр тоже влияет на рейтингование ИР.

Если некоторые из найденных ИР ранее были предложены другим пользователям МАИПС, то по желанию того пользователя, который предоставляет запрос, их оценки могут учитываться либо непосредственно, либо с учётом таких факторов, как степень подобия между этими пользователями и их рейтинг в данной ПрО, который вычисляется как по их собственным оценкам и сведениям о них из социальных сетей, так и по статистике, накопленной МАИПС.

5. *Обработка запросов* для повышения эффективности поиска. Кроме непосредственного выполнения запросов МАИПС определяет уровень компетентности пользователей для различных ПрО, который базируется на таких параметрах, как:

- количество запросов пользователя, основанных на данной онтологии;
- собственная оценка пользователем своей осведомленности в интересующих его ПрО;
- выбранный пользователем уровень читабельности текстовых ИР для данной ПрО - как заданный явно, так и средний для избранных им ИР по запросам, базирующимся на данной онтологии (этот параметр является наиболее объективным);

- релевантность собственных публикаций пользователя этой ПрО, которая определяется путем сопоставления тезауруса, построенного по онтологии ПрО, с тезаурусами публикаций;
- рейтинг пользователя в социальных сетях;
- количество других пользователей МАИПС, выбирающих пользователя за эксперта для поиска рекомендаций в данной ПрО.

б. *Создание рекомендаций* по результатам обработки. В отличие от большинства существующих РС, МАИПС позволяет пользователю явно, непосредственно и динамично управлять средствами создания рекомендаций. Пользователь может учитывать оценки:

- всего сообщества пользователей МАИПС;
- подмножеств пользователей, запросы которых базируются на тех же онтологиях;
- тех пользователей, которые используют наиболее подобные тезаурусы и ключевые слова для запросов;
- запросов с выбранной пользователем темы, в которую могут входить как только собственные запросы пользователя, так и запросы различных пользователей с различными онтологиями и тезаурусами;
- явно указанного подмножества пользователей МАИПС;
- подмножества пользователей МАИПС, построенного по введенным пользователем формальным условиям (например, по месту жительства или по возрасту);
- подмножества запросов самого пользователя, отвечающих определенным условиям (например, построенные в указанный интервал времени и по использованию ключевого слова).

Такие условия пользователь может предоставлять как для каждого запроса по отдельности, так и для определенной группы своих запросов. Результаты построения рекомендаций влияют на упорядочение результатов запроса и позволяют рекомендовать пользователю те ИР, которые не встречаются в результатах его собственных запросов, но считаются МАИПС интересными для него.

7. *Автоматизированное пополнение* профиля МАИПС функционирует в открытой информационной среде Web и поэтому нуждается в динамическом обновлении своих знаний об этой среде. Пути для этого являются проактивный поиск новых сведений о своих пользователях в Web и автоматизированное обновление онтологий ПрО (например, путем обработки семантических вики-ресурсов, экспорта внешних онтологий).

9 Использование интенциональных отношений для моделирования поведения программного агента пользователя ИПС

Оценить эффективность ИПС и её способность удовлетворять различные ИП пользователя достаточно сложно, так как оценки пользователей достаточно субъективны, а самим пользователям обычно сложно строго указать параметры, на основании которых одни ИО оказались предпочтительнее других при достаточно схожих формальных параметрах. Поэтому возникает необходимость в формальном математическом аппарате, позволяющем описывать ИП пользователя и прогнозировать поведение интеллектуального программного агента, представляющего перед ИПС интересы этого пользователя.

Программные агенты (ПА) – современная парадигма программирования, которая позволяет перейти на новый, более интеллектуальный уровень взаимодействия пользователя с программным и аппаратным обеспечением [15]. Она обеспечивает повышение эффективности работы и позволяет пользователям поручить ИС выполнение довольно сложных задач – например, поиск знаний.

Обычно человеческое поведение прогнозируется и объясняется через такие атрибуты отношений, как убеждение, желание, надежды, опасения и т.п., которые называются интенциональными понятиями. Философ Д. Деннет ввел термин интенциональных систем для описания сущностей, поведение которых прогнозируется путём приписывания им атрибутов убеждения, желания и рациональности [16], а Маккарти рассмотрел область применимости таких систем [17]. Чем меньше известно о системе и её структуре, тем более полезны интенциональные объяснения её поведения. Кроме того, для достаточно сложной системы (даже при наличии полной информации о ней) интенциональное объяснение ее поведения часто более практично, чем механистическое.

Интеллектуальный агент, как система, последовательно описывается через интенциональные состояния [18] информационными отношениями (убеждение, знание) и пред-отношениями (желание, намерение, обязательство, совершение, выбор, цель и т.д.). Информационные отношения относятся к информации, которую имеет агент о мире, в котором он существует (в данном случае – о пользователе, интересующей его Про и о нужных ему ИО, а также об информационном пространстве Web, в котором осуществляется процесс поиска информации), тогда как пред-отношения – это то, что некоторым образом руководит действиями агента (сведения о задаче, которую решает пользователь, и о его представлениях о том, что ему надо для её решения).

Убеждения (beliefs) агента выражают его мнение о текущем состоянии мира и о правдоподобии образа действия, приводящего к определенному эффекту.

Желания (desire) описывают предпочтения агента относительно будущих состояний мира или образа действий. Важной чертой желания является то, что агент может иметь несовместимые и недостижимые желания.

Цель (goals) - это непротиворечивое подмножество желаний агента.

Намерения (intension) - непротиворечивое подмножество целей, достижимых ограниченным в ресурсах агентом, и способ их достижения.

Пред- и информационные отношения тесно связаны, поскольку агенты могут делать рациональный выбор, формировать намерения и т.д. на основе информации, имеющейся у них о мире.

Многие интенциональные понятия – такие, как убеждение и желание – являются референциально непрозрачными (например, истинностное значение высказывания «А полагает X» зависит не только от истинностного значения X, но и от А) и поэтому классическая логика в её стандартной форме непригодна для их описания.

Вопрос, какая именно комбинация информационных отношений и пред-отношений лучше всего подходит для характеристики рациональных агентов, является ещё предметом обсуждения. В настоящее время наиболее популярный подход состоит в использовании различных комбинаций убеждений, желаний, целей и намерений. Например, Кохен и Левескье [19] используют два базовых отношения: убеждения и цели (другие отношения определяются через них), а Рао и Джоржеф [20] – три: убеждения, желания и намерения (BDI-архитектура). Однако, все эти формализмы фокусировались только на одном аспекте агентов - их информационных отношениях, а реальная теория агента должна быть представлена логической структурой, которая комбинирует его различные компоненты. Полная теория агента, выраженная в логике с этими свойствами, должна определять то, как атрибуты агента соотносятся между собой, в частности, как соотносятся в агентах информационные отношения и пред-отношения.

Для описания агента пользователя ИПС предлагается использовать формальную модель, включающую четыре интенциональных отношения – знания, убеждения, намерения и цели, а также связи между этими отношениями.

Рассмотрим утверждения типа «А (R)ρ», где А – субъект, ρ – утвердительное предложение, а R – отношение между А и предложением ρ. Если в качестве отношения R рассмотреть убеждение, то утверждение, что «А полагает ρ» истинно, означает, что ρ признается истинным для А независимо от действительной истинности ρ. Предполагается, что существует некто, утверждающий истинность всего предложения «А полагает ρ», т.е. внешний субъект К. Следовательно, существуют две точки зрения: А – внутреннего и К – внешнего субъектов, т.е. в семантике возможных миров [21] необходимо рассматривать два множества возможных миров.

Зафиксируем базисное множество W, отражающее истинную картину и соответствующее знаниям внешнего субъекта К. В W оцениваются и немодальные, и модальные предложения. U – вспомогательное множество возможных миров, представляющее познания внутреннего субъекта А. В мирах $u \in U$ оцениваются только немодальные предложения, выражающие суждения внутреннего субъекта об объектах. Мир $u \in U$ описывает мир $\omega \in W$ с точки зрения А.

В качестве общезначимого критерия отличия знания от убеждения берётся фиксированная точка зрения, представленная базовой системой возможных миров W. В таком случае предложение «А полагает ρ» будет истинно в мире $\omega \in W$ тогда и только тогда, когда ρ истинно в соотнесённом с миром $\omega \in W$ мире $u \in U$. Предложение «А знает ρ» истинно в ω тогда и только тогда, когда ρ истинно в мире ω и в соотнесённом с ω мире $u \in U$. Последнее означает, что субъект утверждает истинность ρ, и ρ действительно истинно.

ω – классический возможный мир, т.е. в нём выполняются законы классической логики, и он полностью определён: любое правильно построенное высказывание языка в нём либо истинно, либо ложно. Но субъект А не всегда располагает полной и неискажённой информацией о мире, поэтому его представления о мире складываются из догадок и предположений. Поэтому для описания миров из U истинностных значений «истинно» и «ложно» явно недостаточно (например, существуют утверждения, истинностная оценка которых неизвестна, но субъект уверен в возможности либо невозможности такой оценки).

Существует зависимость между истинностными значениями и личностными познавательными модальностями (таблица 1).

Таблица 1 – Зависимость между истинностными значениями и личностными познавательными модальностями

Модальность	ω -мир	u-мир
А знает ρ	ρ истинно	ρ истинно
А полагает ρ	ρ произвольно	ρ истинно
А заблуждается в ρ	ρ ложно	ρ истинно
А сомневается в ρ	ρ произвольно	ρ истинно или ρ ложно
А не знает ρ	ρ произвольно	ρ произвольно

Введем модальные операторы B, D и I.

$B\rho$ означает «некто полагает ρ». В случае N познающих субъектов можно ввести индекс $i, i = \overline{1, N}$, $B_i\rho$ означает «i-й субъект полагает ρ».

\mathfrak{L} – язык классической пропозициональной логики. \mathfrak{L}^e – эпистемический язык, который содержит счётное множество пропозициональных переменных $Var = \{\rho, q, r, \dots\}$, связки & и \neg , скобки и семейство обобщённых модальных операторов $\{B_i, D_i, I_i; i, i = \overline{1, N}\}$. F – множество формул языка \mathfrak{L} , такое, что:

1. Если $\alpha \in Var$, то $\alpha \in F$.

2. Если $\alpha \in F$, то $\neg\alpha \in F$.

3. Если $\alpha, \beta \in F$, то $(\alpha \& \beta) \in F$.

Множество формул языка \mathfrak{F}^e – такое множество F^e , что:

1. Если $\alpha \in F$, то $\alpha \in F^e$.

2. Если $\alpha \in F^e$, то $B_i\alpha \in F^e$, $i = \overline{1, N}$.

3. Если $\alpha \in F^e$, то $D_i\alpha \in F^e$, $i = \overline{1, N}$.

4. Если $\alpha \in F^e$, то $I_i\alpha \in F^e$, $i = \overline{1, N}$.

Остальные связки (\vee, \supset, \equiv) определяются обычным образом. Множество F^e будем называть базовым, а множество F – вспомогательным.

Для языка \mathfrak{F}^e предлагается следующая семантика возможных миров

$\wp = \langle W, U, s, \varphi \rangle$ – модель языка \mathfrak{F}^e тогда и только тогда, когда:

1. W – непустое базовое множество возможных миров;

2. U – непустое вспомогательное множество возможных миров;

3. s – функция типа: $s: W \rightarrow U$;

4. φ – функция приписывания.

$\varphi(\rho) = \langle TW(\rho), TU(\rho), FU(\rho) \rangle \quad \forall \rho \in \text{Var}$, где $TW(\rho) \subseteq W$ и $TU(\rho) \subseteq U$ – множество миров,

где ρ истинно, а $FU(\rho) \subseteq U$ – множество миров, где ρ ложно; причем $TU(\rho) \cap FU(\rho) = \emptyset$, но в системе миров U допускаются неизвестные значения истинности ρ .

Пусть α и β принадлежат вспомогательному множеству формул F , γ и δ – формулы базового множества F^e , $i = \overline{1, N}$. Следующие схемы аксиом и правило вывода описывают модальный оператор убеждения:

A1. $B_i(\alpha \& \beta) \equiv (B_i\alpha \& B_i\beta)$;

A2. $B_i(\alpha \vee \beta) \equiv (B_i\alpha \vee B_i\beta)$;

A3. $B_i(\alpha \rightarrow \beta) \equiv (B_i\alpha \rightarrow B_i\beta)$;

A4. $B_i\neg(\alpha \& \beta) \equiv (B_i\neg\alpha \vee B_i\neg\beta)$;

A5. $B_i\neg(\alpha \vee \beta) \equiv (B_i\neg\alpha \& B_i\neg\beta)$;

A6. $B_i\neg(\alpha \rightarrow \beta) \equiv (B_i\alpha \& B_i\neg\beta)$;

A7. $B_i\alpha \rightarrow \neg B_i\neg\alpha$;

A8. $B_i\alpha \rightarrow B_i\neg\neg\alpha$;

A9. $B_i\neg\neg\alpha \rightarrow B_i\alpha$;

Правило вывода «modus ponens»: если $\gamma, \gamma \rightarrow \delta$, то δ .

Введем в язык \mathfrak{F}^e новые операторы, характеризующие познания: оператор знания K и оператор заблуждения E . $K_i\alpha$ означает « i -тый субъект знает α », а $E_i\alpha$ – « i -тый субъект ошибается, предполагая α ».

A10. $K_i\alpha \rightarrow (B_i\alpha \& \alpha)$;

В схеме A10 использован знак импликации, поскольку если при определении заблуждения достаточно указать на ложность того, что субъект принимает как истинное, то из того факта, что нечто истинно для субъекта и истинно в фиксированной системе, ещё нельзя заключить о том, что субъект знает это: $(B_i\alpha \& \alpha) \& \neg K_i\alpha$.

A11. $E_i\alpha \equiv (B_i\alpha \& \neg\alpha)$.

Имеет место транзитивное отношение достижимости:

A12. $K_i\alpha \rightarrow K_i K_i\alpha$

Эта аксиома описывает процесс исследования индивидуумом своих убеждений: он осознает то, что он знает.

A13. $B_i\alpha \rightarrow K_i B_i\alpha$

Эта аксиома говорит о том, что агент осознает, чего он не знает.

Введём модальный оператор цели. Утверждение, что «А желает ρ » истинно, означает, что ρ признаётся предпочтительным для А независимо от действительных возможностей А и достижимости для него ρ .

Субъект А может иметь несовместимые желания, и он не должен полагать, что его желания достижимы. Поэтому введём более узкое понятие – цель. Цель – это совместимое подмножество желаний. $D_i\rho$ означает «некто имеет цель ρ ». $D_i\rho$ означает « i -й субъект имеет цель ρ ». Следующие схемы аксиом и правило вывода описывают модальный оператор цели:

$$B1. D_i(\alpha \& \beta) \equiv (D_i\alpha \& D_i\beta);$$

$$B2. D_i(\alpha \vee \beta) \equiv (D_i\alpha \vee D_i\beta).$$

Цели (в отличие от желаний) не противоречат друг другу:

$$B3. D_i\neg\alpha \equiv \neg D_i\alpha;$$

$$B4. D_i\neg(\alpha \& \beta) \equiv (D_i\neg\alpha \vee D_i\neg\beta);$$

$$B5. D_i\neg(\alpha \vee \beta) \equiv (D_i\neg\alpha \& D_i\neg\beta);$$

$$B6. D_i\alpha \equiv D_i\neg\neg\alpha;$$

$$B7. D_i(\alpha \rightarrow \beta) \rightarrow (D_i\alpha \rightarrow D_i\beta);$$

$$B8. V_i(\alpha \rightarrow \beta) \& D_i(\alpha) \rightarrow D_i\beta;$$

$$B9. V_i(D_i\alpha) \equiv D_i(\alpha).$$

Утверждение, что «А намерен ρ » истинно, означает, что «А желает ρ » и «А полагает $Y(\rho)$ », где $Y(\rho)$ – способ достигнуть ρ . $I_i\rho$ означает «некто намерен ρ ». $I_i\rho$ означает « i -й субъект намерен ρ ». Следующие схемы аксиом и правило вывода описывают модальный оператор намерения:

$$C1. I_i(\alpha \& \beta) \equiv (I_i\alpha \& I_i\beta);$$

$$C2. I_i(\alpha \vee \beta) \equiv (I_i\alpha \vee I_i\beta).$$

Намерения не противоречат друг другу:

$$C3. \neg I_i(\alpha) \equiv I_i\neg\alpha;$$

$$C4. I_i\neg(\alpha \& \beta) \equiv (I_i\neg\alpha \vee I_i\neg\beta);$$

$$C5. I_i\neg(\alpha \vee \beta) \equiv (I_i\neg\alpha \& I_i\neg\beta);$$

$$C6. I_i\alpha \equiv I_i\neg\neg\alpha;$$

$$C7. I_i(\alpha \rightarrow \beta) \rightarrow (I_i\alpha \rightarrow I_i\beta);$$

$$C8. I_i\alpha \rightarrow D_i\alpha;$$

$$C9. V_i(I_i\alpha) \equiv I_i(\alpha).$$

Очевидно, что предложенный выше набор схем аксиом является непротиворечивым, то есть $\forall \alpha \in F^e$ не выводима $\alpha \& \neg\alpha$ (так как убеждения, цели и намерения каждого конкретного субъекта не противоречат друг другу по определению). В то же время предложенная система аксиом в общем случае не полна, то есть $\exists \alpha \in F^e$, для которых не выводимы ни α , ни $\neg\alpha$ (это вызвано неполнотой знаний и убеждений конкретного субъекта).

Выбор данной модальной системы позволяет моделировать поведение разумного целеустремлённого субъекта, некоторые предположения которого могут оказаться ошибочными, но который обладает совершенной способностью к логической интроспекции относительно того, что он предполагает и чего не предполагает [22]. В частности, в МАИПС таким субъектом является пользователь и представляющий его интересы интеллектуальный ПА. Использование модальностей позволяет смоделировать его поведение при оценке различных ИО.

Выводы

Разработка онтологической модели ИП пользователя при семантическом поиске в Web обеспечивает совместное использование методов выработки рекомендаций, доступа к внешним источникам информации, индуктивного извлечения знаний и технологий Semantic Web, что позволяет более эффективно обеспечить пользователя необходимыми сведениями, а явным образом выбранные методы рекомендаций и онтологические описания Про обеспечивают пользователю понимание поведения такой системы. Применение парадигмы интеллектуальных программных агентов при описании элементов такой поисковой системы позволяют прогнозировать её поведение.

Список источников

- [1] *Розушина Ю.В.* Управление знаниями на основе онтологий в дистанционном обучении. – LAP LAMBERT Academic Publishing, 2013. – 92 с.
- [2] *Розушина Ю.В.* Знание-ориентированные средства поддержки семантического поиска в Web // Материалы IV международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» OSTIS-2014, Минск, БГУИР, 2014. – С.339-352.
- [3] *Хорошевский В.Ф.* Онтологические модели и Semantic Web: откуда и куда мы идем?// Сб. трудов симпозиума «Онтологическое моделирование». Звенигород, 19–20 мая 2008 г. М., ИПИ РАН, 2008.
- [4] *Amerland D.* Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. – Que Publishing, 2013. – 230 p.
- [5] *Ricci F., Rokach L., Shapira B., Kantor P.* Recommender Systems Handbook. – Springer, 2011. – 842 p.
- [6] *Розушина Ю.В.* Менеджмент знаний в рекомендующих системах на основе онтологий // Сборник трудов XIII международной конф. им. Т.А.Таран "Интеллектуальный анализ информации ИАИ-2013", К., Просвіта, 2013. – С. 14-20.
- [7] *Middleton S., De Roure D., Shadbolt N.* Ontology-Based Recommender Systems // in Handbook on Ontologies, Edt. by S.Staab, R.Studer, Springer, 2009. – P. 779-796.
- [8] *Браславский П.И., Гольдштейн С.Л., Ткаченко Т.Я.* Тезаурус как средство описания систем знаний // Информационные процессы и системы. – 1997. – № 11, Серия 2. – С. 16-22.
- [9] *Лукашевич Н.В.* Тезаурусы в задачах информационного поиска. – М.: Издательство Московского университета, 2011. – 512 с.
- [10] *Гладун А.Я., Розушина Ю.В.* Онтологии и мультилингвистические тезаурусы как основа семантического поиска информационных ресурсов Интернет // The Proc. of XII-th Intern. Conf. KDS'2006, Varna, Bulgaria. – P. 115-121.
- [11] *Quinlan J.R.* Discovery rules from large collections of examples: a case study // Expert Systems in the Microelectronic Age. – Edinburg, 1979. – P. 87-102.
- [12] *Rogushina J., Gladun A.* Ontology-based competency analyses in new research domains // Journal of Computing and Information Technology. V.20, N. 4, 2012. – P.277-293.
- [13] *Розушина Ю.В.* Использование критериев оценки удобочитаемости текста для поиска информации, соответствующей реальным потребностям пользователя // Проблемы программирования. – 2007. № 3. – С. 76-87.
- [14] *Розушина Ю.В., Гладун А.Я.* Репозитории онтологий как средство повторного использования знаний для распознавания информационных объектов // Онтология проектирования, № 1 (7), 2013. – С.35-50.
- [15] *Плескач В.Л., Розушина Ю.В.* Агентні технології. – Монографія. К.: Київ. нац. торг.-екон. ун-т, 2005. – 338 с.
- [16] *Dennett D.C.* The Intensional Stance. The MIT Press: Cambridge, MA, 1987. – 282 p.
- [17] *McCarthy J.* Ascribing mental qualities to machines. Technical report, Stanford University AI Lab., Stanford, CA 94305, 1978. – 84 p.
- [18] *Konolige K.* A Deduction Model of Belief. Pitman Publishing: London and Morgan Kaufmann: San Mateo, CA, 1986. – 226 p.
- [19] *Cohen P.R., Levesque H.J.* Intention is choice with commitment // Artificial Intelligence, 42, 1990. – P.213-261
- [20] *Rao A.S., Georgeff M.P.* Modeling rational agents within a BDI-architecture // In R. Pikes and E. Sandewall, eds.. Proc. of Knowledge Representation and Reasoning (KR&R-91), Morgan Kaufmann Publishers: San Mateo, CA, April 1991. – P. 473-484.
- [21] *Hintikka J.* Knowledge and Belief. Cornell University Press: Ithaca, NY, 1962. – 124 p.

- [22] *Рогушина Ю.В.* Программные агенты: определения, таксономии, модели // Управляющие системы и машины, 2001, N 5. - С.39-45

DESIGN OF THE ONTOLOGICAL MODEL OF USER'S INFORMATIONAL NEED IN SEMANTIC SEARCH

J.Rogushina

*Institute of software systems of National academy of sciences of Ukraine , Kiev, Ukraine
ladamandraka2010@gmail.com*

Abstract

Design of the ontological model of user's informational need that is satisfied by means of semantic search is analyzed. The structure of this model (it's main components – users and user's personal program agents, informational resources, domain ontologies, informational objects etc. and relations between these components), means of model refinement and method of its use for semantic search are proposed. Reasonability of ontological approach for development of this model that represent personal (about users) and collaborative (about groups of users) knowledge is grounded. Paradigm of intelligent software agents is used for description of retrieval system behavior forecasting.

Key words: *semantic search, ontological model, thesaurus, recommending system, software agent.*

References

- [1] *Rogushina J.V.* Upravlenie znaniyami na osnove ontologij v distantsionnom obuchenii. [Ontology-based knowledge management in distance learning.] – LAP LAMBERT Academic Publishing, 2013.– 92 p. (In Russian).
- [2] *Rogushina J.V.* Znanie-orientirovannye sredstva podderzhki semanticheskogo poiska v Web [Knowledge-based support tools for semantic Web search] // Proceedings of the IVth international scientific conference OSTIS-2014, Minsk, BGUIR, 2014. – p.339-352. (In Russian).
- [3] *Khoroshevskij V.F.* Ontologicheskie modeli i Semantic Web: otкуда i kuda my idem? [Ontological models and Semantic Web: how and where are we going?]: Proceedings of the “Ontological modeling” symposium. Zvenigorod, 19–20 May 2008 г. Moscow., IPI RAS, 2008. (In Russian)
- [4] *Amerland D.* Google Semantic Search: Search Engine Optimization (SEO) Techniques That Gets Your Company More Traffic, Increases Brand Impact and Amplifies Your Online Presence. – Que Publishing, 2013. – 230 p.
- [5] *Ricci F., Rokach L., Shapira B., Kantor P.* Recommender Systems Handbook. – Springer, 2011. – 842 p.
- [6] *Rogushina J.V.* Menedzhment znaniy v rekomenduyushhikh sistemakh na osnove ontologij [Knowledge Management in recommender systems based on ontologies] // Proceedings of the XIIIth international conference "Intellectual information analysis IAI-2013", Kiev, Prosvita, 2013. – pp. 14-20. (In Russian)
- [7] *Middleton S., De Roure D., Shadbolt N.* Ontology-Based Recommender Systems // in Handbook on Ontologies, Edt. by S.Staab, R.Studer, Springer, 2009. – P. 779-796.
- [8] *Braslavskij P.I., Gol'dshtejn S.L., Tkachenko T.Ya.* Tezaurus kak sredstvo opisaniya sistem znaniy [Thesaurus as a means of describing knowledge systems] // Information processes and systems. – 1997. – No. 11, Series 2. – pp. 16-22. (In Russian)
- [9] *Lukashevich N.V.* Tezaurusy v zadachakh informatsionnogo poiska [Thesauri in information retrieval tasks]. – Moscow: Moscow university publishing, 2011. – 512 p. (In Russian)
- [10] *Gladun A.Ya., Rogushina Yu.V.* Ontologii i multilingvisticheskie tezaurusy kak osnova semanticheskogo poiska informatsionnykh resursov Internet [Thesauri and ontologies as a basis for multilingvistic semantic search for Internet information resources] // The Proc. of XII-th Intern. Conf. KDS'2006, Varna, Bulgaria. – P. 115-121.
- [11] *Quinlan J.R.* Discovery rules from large collections of examples: a case study // Expert Systems in the Microelectronic Age. – Edinburg, 1979. – P. 87-102.
- [12] *Rogushina J., Gladun A.* Ontology-based competency analyses in new research domains // Journal of Computing and Information Technology. V.20, N. 4, 2012. – P.277-293.
- [13] *Rogushina J.V.* Ispol'zovanie kriteriev otsenki udobochitaemosti teksta dlya poiska informatsii, sootvetstvuyushhej re-al'nym potrebnyostyam pol'zovatelya [Using readability criteria evaluation to search for information relevant to the needs of the real user] // Problems of programming. – 2007. –No 3. – pp. 76-87. (In Russian)

- [14] **Rogushina J., Gladun A.** Repozitorii ontologij kak sredstvo povtornogo ispol'zovaniya znaniy dlya raspoznavaniya informatsionnykh ob"ektov [Ontology repository as a means of re-use of knowledge for the recognition of information objects] // *Ontology of designing*, No. 1 (7), 2013. – pp.35-50. (In Russian)
- [15] **Pleskach V.L., Rogushina J.V.** Agent technologies. – Monograph. Kiev: Kiev national trading institute, 2005. – 338 p. (In Ukrainian)
- [16] **Dennett D.C.** *The Intensional Stance*. The MIT Press: Cambridge, MA, 1987. – 282 p.
- [17] **McCarthy J.** Ascribing mental qualities to machines. Technical report, Stanford University AI Lab., Stanford, CA 94305, 1978. – 84 p.
- [18] **Konolige K.** *A Deduction Model of Belief*. Pitman Publishing: London and Morgan Kaufmann: San Mateo, CA, 1986. – 226 p.
- [19] **Cohen P.R., Levesque H.J.** Intention is choice with commitment // *Artificial Intelligence*, 42, 1990. – P.213-261
- [20] **Rao A.S., Georgeff M.P.** Modeling rational agents within a BDI-architecture // In R. Pikes and E. Sandewall, eds.. *Proc. of Knowledge Representation and Reasoning (KR&R-91)*, Morgan Kaufmann Publishers: San Mateo, CA, April 1991. – P. 473-484.
- [21] **Hintikka J.** *Knowledge and Belief*. Cornell University Press: Ithaca, NY, 1962. – 124 p.
- [22] **Rogushina J.V.** Programmnye agenty: opredeleniya, taksonomii, modeli [Software agents: definitions, taxonomy model] // *Upravlyayushhie sistemy i mashiny*, 2001, No 5. - pp.39-45 (In Russian)
-

Сведения об авторах



Рогущина Юлия Витальевна, 1967 г. рождения. Окончила Киевский государственный университет им. Т.Г.Шевченко в 1989 году. Степень кандидата физико-математических наук получила в Институте кибернетики им. В.М. Глушкова (г. Киев, Украина) в 1995 году. В настоящее время работает старшим научным сотрудником Института программных систем, Национальной академии наук Украины. Ее исследовательские интересы включают разработку интеллектуальных информационных систем; теорию поведения программных агентов, индуктивное извлечение знаний, семантический поиск в Web, онтологический анализ, технологий Semantic Web.

Она является автором более 150 публикаций, в том числе монографии «Агентные технологии».

Rogushina Julia Vitalievna (b. 1967). She received the M.Sc. from Kiev Taras Shevchenko State University in 1989. She received her PhD degree in Computer Science in Glushkov`s Institute of Cybernetics, Kiev, in 1995. She is a senior researcher at the Institute of Software Systems, National Academy of Sciences of Ukraine. Her research interests include the development and application of intelligent information systems; theory of software agents behavior, inductive knowledge acquisition, intelligent information retrieval, ontological analysis, Semantic Web technologies. She has more than 150 publications in scientific journals and conferences, monograph «Agent technologies».