

УДК 004.822:004.89

СЕМАНТИЧЕСКАЯ ТЕХНОЛОГИЯ РАЗРАБОТКИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ, ОРИЕНТИРОВАННАЯ НА ЭКСПЕРТОВ ПРЕДМЕТНОЙ ОБЛАСТИ

Ю.А. Загорулько

*Институт систем информатики им. А.П. Ершова
Сибирского отделения Российской академии наук, Новосибирск, Россия
zagor@iis.nsk.su*

Аннотация

В статье представлена технология построения интеллектуальных систем, ориентированная на экспертов, т.е. специалистов в предметных областях, для которых разрабатываются системы. В качестве базиса представления знаний предлагается модель, интегрирующая взаимодополняющие друг друга методы и средства представления и обработки знаний следующих моделей: онтологической, сетевой, продукционной, недоопределённой и императивной при ведущей роли онтологической модели. Инструментальные средства, построенные на основе интегрированной модели, могут применяться для разработки широкого класса интеллектуальных систем. Для того чтобы сделать эти средства доступными для непосредственного использования экспертами, была предложена технология построения интеллектуальных систем, базирующаяся на специализированных оболочках, настроенных на определенные области знаний и классы задач. При этом сами специализированные оболочки могут создаваться инженерами знаний путем частичного заполнения баз знаний соответствующих универсальных оболочек с использованием включенных в них средств описания предметных знаний.

Ключевые слова: интегрированная модель представления знаний, онтология, интеллектуальная система, портал знаний, специализированная программная оболочка.

Введение

В связи с переходом к экономике знаний, когда основным фактором, определяющим развитие общества, становятся накопленные человечеством знания и умения, их доступность широкому кругу пользователей, всё более востребованными становятся интеллектуальные системы (ИС), особенно такая их разновидность, как системы, основанные на знаниях (СОЗ), и порталы знаний. Соответственно растут и потребности в эргономичных и эффективных средствах разработки такого класса систем.

Наиболее трудоемким и в то же время самым ответственным этапом разработки интеллектуальной системы является построение её базы знаний (БЗ) [1], которая, в конечном счете, определяет полезность и качество всей системы. В связи с этим определяющую роль играют модели и методы представления и обработки знаний, а также построенные на их основе инструментальные средства и технологии создания ИС.

В рамках работ по искусственному интеллекту накоплен широкий спектр моделей, методов и средств представления и обработки знаний. Сетевая модель предлагает для представления знаний такие средства, как семантические сети [2] и фреймы [3]. Первые используются в качестве универсальной памяти для хранения любой информации, которую можно представить в терминах объектов и отношений между ними. Фреймы, представляющие объектно-ориентированный подход в искусственном интеллекте, служат как для повышения уровня

представления знаний, так и для обеспечения возможности совместного использования декларативных и процедурных знаний. Продукционная модель [4] позволяет представлять знания экспертов в виде множества правил вида «ЕСЛИ-ТО» и организовывать на них логический вывод. В рамках парадигмы программирования в ограничениях [5] знания могут задаваться в виде системы ограничений на значения параметров моделируемых объектов. Недоопределённая модель [6], объединяющая методы программирования в ограничениях с аппаратом недоопределённых типов данных [7], даёт возможность оперировать неточно заданными значениями и частично описанными объектами. Широкую популярность у разработчиков баз знаний получили онтологии [8–10], с помощью которых можно представить согласованную систему понятий моделируемой области знаний. В последнее время для разработки онтологий и баз знаний успешно применяются дескриптивные логики [11] и построенные на их основе языки и системы представления знаний.

На основе описанных выше моделей и методов созданы инструментальные системы, предназначенные для построения интеллектуальных систем и их баз знаний. Однако у этих систем имеется один важный недостаток – они, как правило, базируются только на одной модели представления знаний и снабжены каким-то одним из упомянутых выше средств. В то время как при создании прикладных ИС часто возникает необходимость их комплексного использования, поскольку ни одно из них, взятое в отдельности, не может обеспечить в полном объёме потребности разработки реальной прикладной системы. В связи с этим инструментальные средства разработки ИС должны включать различные взаимодополняющие друг друга методы и средства представления и обработки знаний.

Другим важным требованием к инструментальным средствам и построенным на их основе технологиям является то, что возможности представления и обработки знаний должны обладать высокой выразительной силой и гибкостью, быть понятными и простыми в использовании как для инженеров знаний – специалистов по представлению знаний, так и для экспертов – специалистов в конкретных предметных областях (ПО).

Такой эффект может быть достигнут только тогда, когда все необходимые методы и средства представления знаний будут объединены в одном формализме (модели и/или языке). Такая интеграция не только создаёт предпосылки для построения мощных баз знаний (за счёт совместного использования различных взаимодополняющих друг друга методов и средств), но и обеспечивает концептуальное единство создаваемых баз знаний (все компоненты БЗ становятся связанными общими понятиями).

На основе такого формализма могут быть созданы инструментальные средства, с помощью которых эксперты смогут сами строить базу знаний ИС в полном объёме, не прибегая к помощи посредников – инженеров знаний и программистов. Благодаря этому значительно упрощается разработка базы знаний ИС и повышается её качество. Кроме того, интеграция всех требуемых средств в рамках одной модели (языка) создаёт предпосылки для разработки на её основе эффективных прикладных ИС, так как устраняются неизбежные расходы ресурсов на организацию взаимодействия программных систем, реализующих отдельные методы и средства представления знаний.

В статье рассматриваются интегрированная модель представления знаний, объединяющие методы и средства нескольких различных моделей, построенные на основе этой модели инструментальные средства и технологии разработки ИС, ориентированные на экспертов.

1 Интегрированная модель представления знаний

1.1 Требования к модели представления знаний со стороны интеллектуальной системы

Вначале рассмотрим основные требования к интегрированной модели представления знаний (ИМПЗ-модели).

Для того чтобы модель представления знаний была пригодна для разработки широкого класса интеллектуальных систем, она должна обеспечивать представление всех видов знаний, необходимых для функционирования ИС.

Прежде всего, в базе знаний ИС необходимо представлять знания о ПО, в рамках которой будет функционировать система (см. рисунок 1). К ним относятся знания об основных сущностях (понятиях и объектах) ПО, а также знания о том, каким образом эти сущности связаны между собой. К последним относятся знания об отношениях, непосредственно связывающих понятия, а также более сложные виды знаний, представляющие различного рода зависимости между понятиями ПО, как логические, так и функциональные. В частности, при разработке большого класса ИС возникает необходимость в представлении знаний о ПО в виде правил «ЕСЛИ-ТО». Также довольно часто возникает потребность в представлении знаний в виде ограничений на значения параметров, описывающих некоторый объект или систему объектов. Для лучшего структурирования ПО и обеспечения более лаконичного её описания важно иметь возможность выстраивания понятий ПО в иерархию «общее-частное» и поддержки наследования свойств по этой иерархии.

Базы знаний большинства ИС также содержат конкретные знания из ПО (предметные знания), которые представляются в виде экземпляров понятий (конкретных объектов) и связей между ними – экземпляров отношений или ограничений, заданных на значениях атрибутов экземпляров понятий.

Другим важным типом знаний, которые нужно представлять в ИС, являются знания о задачах и способах их решения в рамках моделируемой ПО (методах и алгоритмах). Эти знания характеризуют проблемную область (ПрО) интеллектуальной системы. Такого рода знания могут быть как декларативными, так и процедурными. К первому типу относятся знания, описывающие пространство решаемых ИС задач, включая разбиение задач на подзадачи и связь подзадач с методами их решения, ко второму – знания, представляющие как методы решения задач, так и конкретные алгоритмы.

Кроме знаний, представленных в базе знаний, практически в любых ИС необходимо представлять знания, описывающие фрагмент действительности (ситуацию), который задает контекст и входные данные для задач, решаемых ИС. Такие знания, как и предметные знания, обычно представляются в виде экземпляров понятий и отношений и/или ограничений, заданных на значениях атрибутов экземпляров понятий.

При описании предметных знаний, контекста и входных данных задачи довольно часто возникает необходимость представлять знания об объектах, имеющих неточные значения атрибутов, путем задания оценок таких значений, например, в виде множеств возможных значений и/или ограничений на значения. Для представления таких знаний необходимо обеспечить возможность использования в экземплярах понятий атрибутов, имеющих недоопределённые значения, и задания экземпляров ограничений, связывающих такие атрибуты. Наличие указанного вида знаний создаёт потенциальную возможность уточнения значений таких атрибутов через механизм удовлетворения ограничений [5].

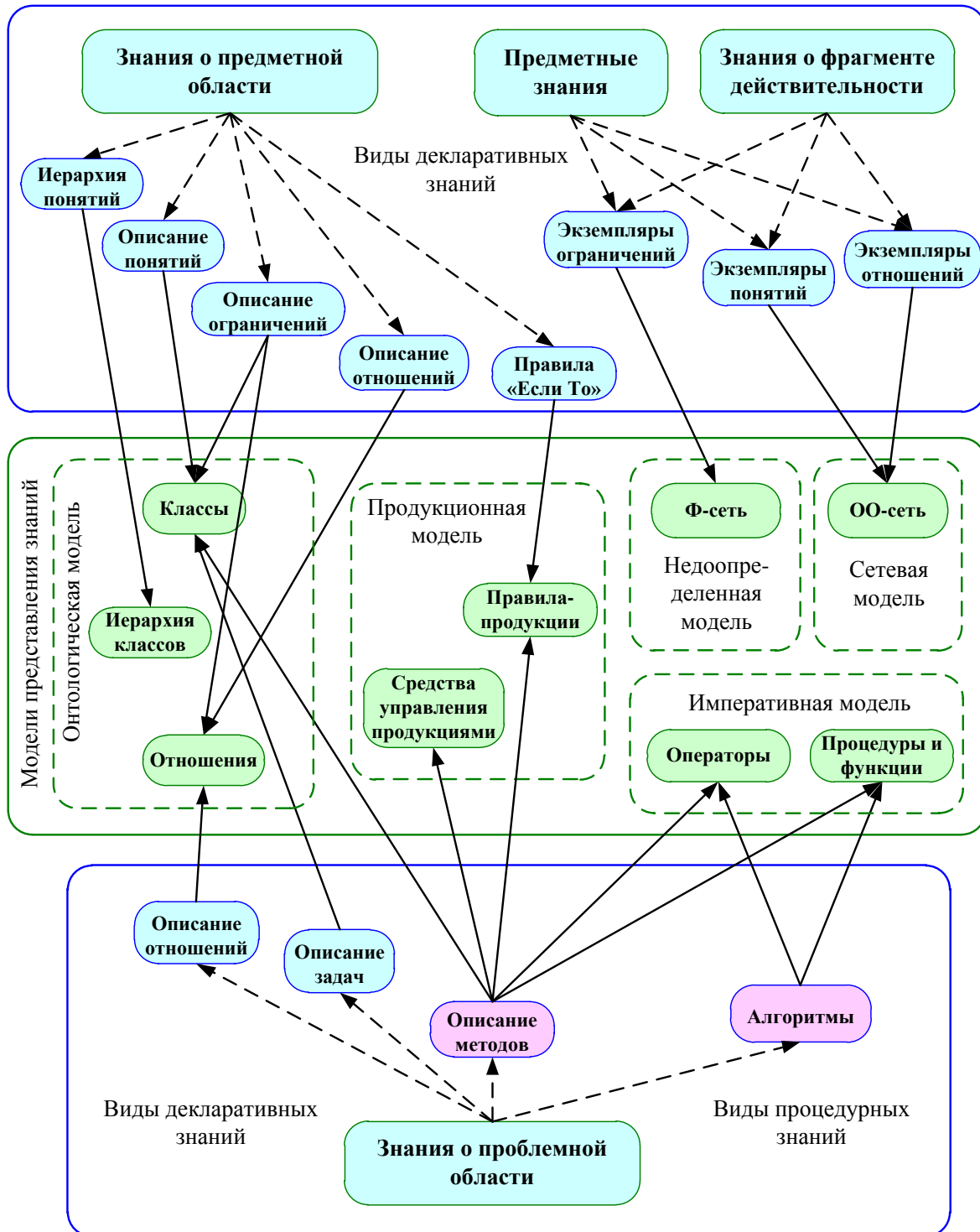


Рисунок 1 – Виды знаний и возможные способы их представления

Рассмотренные выше виды знаний, необходимые для полноценного функционирования ИС, не могут быть представлены средствами одной модели. Для этого требуются средства нескольких моделей представления знаний. Так, для представления знаний о предметной области требуется задействовать средства двух моделей – онтологической и продукционной (см. рисунок 1). Часть предметных знаний и знаний о фрагменте действительности может быть представлена средствами сетевой модели (в объектно-ориентированной семантической

сети), другая часть – средствами недоопределённой модели (в функциональной сети). Для представления знаний о проблемной области требуются средства трёх моделей – онтологической, продукционной и императивной.

Таким образом, возникает необходимость в модели представления знаний, которая обладает набором методов и средств, достаточным для представления в ИС всех необходимых знаний. Прежде всего, такая модель должна обеспечивать представление как декларативных, так и процедурных знаний, что является необходимым требованием, так как без представления этих двух типов знаний невозможна разработка полноценной ИС. Кроме того, она должна обеспечивать представление и обработку как точных, так и неточных (недоопределённых) данных, что делает её пригодной для решения задач в условиях неопределённости знаний и данных.

1.2 Концепция и состав интегрированной модели представления знаний

Исходя из перечисленных выше требований, предложена концепция интегрированной модели представления знаний, которая объединяет различные взаимодополняющие друг друга средства и методы представления и обработки знаний следующих моделей: онтологической, сетевой, продукционной, императивной и недоопределённой. (Последняя включает аппарат недоопределённых типов данных и разработанный на его основе метод недоопределённых вычислительных моделей [7].) Выше было показано, что комбинация методов и средств представления знаний, предоставляемых этими моделями, является достаточной для представления в базе знаний ИС всех необходимых знаний.

Все средства представления и обработки знаний интегрированной модели объединены в рамках одного формализма. Такой подход обеспечивает возможность совместного использования всех средств интегрируемых моделей, а, следовательно, и совместного описания декларативных и процедурных знаний.

Формально ИМПЗ-модель может быть представлена шестеркой:

$$M_{IKRM} = \langle ONT, SN, FN, PS, IM, GM \rangle,$$

где:

- *ONT* – прикладная онтология, задающая описания основных сущностей (понятий и отношений) предметной (ПО) и проблемной областей (ПрО) приложения в виде классов объектов и отношений;
- *SN* – объектно-ориентированная семантическая сеть (ОО-сеть), свойства объектов и отношений которой определяются онтологией *ONT*;
- *FN* – функциональная сеть, включающая экземпляры ограничений, заданные на значениях атрибутов объектов семантической сети *SN*;
- *PS* – система правил-продукций, описанных в терминах классов и отношений онтологии *ONT* и операций над их экземплярами;
- *IM* – средства императивного программирования;
- *GM* – общий механизм вывода/обработки знаний и данных, представленных в ИМПЗ-модели, объединяющий механизмы вывода/обработки каждой из интегрируемых моделей.

Прикладная онтология *ONT* составляет базис интегрированной модели представления знаний, так как она определяет главные сущности моделируемых предметной и проблемной областей, которые используются всеми остальными средствами модели.

Дадим формальное описание прикладной онтологии:

$$ONT = \langle C, R_{In}, R, T, D, A_C, A_R, M_C, M_R, Cntr, P_R, R_{rev} \rangle,$$

где:

- $C = \{C_1, \dots, C_m\}$ – конечное непустое множество классов, описывающих понятия некоторой предметной или проблемной области;
- R_{In} , где $R_{In} \subseteq C \times C$ – строгий частичный порядок на множестве классов C , задающий отношение наследования;
- $R = \{R_1, \dots, R_n\}$, $R_i \subseteq C_j \times C_k$, $C_j \in C$, $C_k \in C$ – конечное множество бинарных отношений, заданных на классах C ;
- $T = \{T_1, \dots, T_t\}$ – множество типов данных;
- $D = \{d_1, \dots, d_p\}$ – множество доменов, таких что $d_i = \{s_1, \dots, s_k\}$, где s_i – значение некоторого простого типа T_j из T ;
- $A_C = \{a_1, \dots, a_r\}$, $a_j \subseteq C_k \times (T \cup D)$, $C_k \in C$ – конечное множество атрибутов, описывающих свойства классов C ;
- $A_R = \{a_1, \dots, a_v\}$, $a_j \subseteq R_k \times (T \cup D)$, $R_k \in R$ – конечное множество атрибутов, описывающих дополнительные свойства отношений R ;
- M_C , где $M_C : C \rightarrow 2^{A_C}$ – отображение, задающее связь классов C с атрибутами A_C ;
- M_R , где $M_R : R \rightarrow 2^{A_R}$ – отображение, задающее связь отношений R с атрибутами A_R ;
- $Cntr$ – множество ограничений, заданных на атрибутах классов, т.е. логических выражений вида $Exp_i(e_{i1}, \dots, e_{iw})$, где $e_{ik} \in A_i$ или $e_{ik} \in T_j$, т.е. e_{ik} либо имя атрибута, либо константа;
- P_R – множество математических свойств (симметричность, рефлексивность, транзитивность, антирефлексивность, антисимметричность)¹, которые могут быть заданы для отношений R ;
- R_{rev} , где $R_{rev} \subseteq R \times R$ – множество пар взаимобратных отношений из R .

В онтологии ONT для каждого класса C_i может быть задан свой набор атрибутов $A_{C_i} \subseteq A_C$ и ограничений $Cntr_{C_i} \subseteq Cntr$, связывающих значения атрибутов объектов.

Благодаря включению в ИМПЗ-модель аппарата недоопределённых типов данных, все простые типы данных, входящие в T , имеют недоопределённые расширения, обеспечивающие возможность представления недоопределённых значений (н-значений в виде интервала или множества допустимых значений) и оперирования ими. Это позволяет, в частности, задавать атрибутам объектов ОО-сети н-значения.

С помощью ограничений $Cntr$ могут задаваться функциональные зависимости между атрибутами сущностей не только в описаниях классов объектов, но и отношений. (Заметим, что в описаниях бинарных отношений ограничения задаются на атрибутах их аргументов.) Такие ограничения представляют собой множество логических выражений (формул), связывающих арифметическими, теоретико-множественными и логическими операциями значения атрибутов одного или нескольких объектов. Следует заметить, что благодаря включению в ИМПЗ-модель метода недоопределённых моделей с помощью ограничений задаются не только условия корректности значений связанных ими атрибутов, но и функциональные интерпретации каждого из этих значений относительно других, что обеспечивает потенциальную возможность вычисления (уточнения) неизвестных (неточных) значений атрибутов через известные (или более точные).

¹ Здесь и далее автор оперирует далеко не полным набором свойств симметрии бинарных отношений. Прим. ред.

С помощью отношения R_{In} на классах онтологии может быть задана иерархия «общее-частное», по которой организуется наследование свойств (включая атрибуты, отношения и ограничения) вышестоящих классов нижестоящими, что повышает изобразительные возможности онтологии.

В зависимости от потребностей представления знаний о ПО любому бинарному отношению $R_k \in R$ может быть «приписано» одно или несколько не противоречащих друг другу математических свойств $P_k \in P$ из следующего набора: рефлексивность, симметричность, транзитивность, антирефлексивность, антисимметричность, а также задано обратное отношение.

Таким образом, прикладная онтология ONT , задавая комплексное описание понятий и отношений моделируемых ПО и ПрО в виде классов объектов и отношений, инкапсулирующих в себе семантические свойства и ограничения на свои атрибуты и аргументы, тем самым определяет структуру семантической и функциональной сети, с помощью которых представляются предметные знания и знания о задачах. Онтология также задаёт множество сущностей, в терминах которых описываются продукционные правила, реализующие процессы вывода и обработки информации.

Покажем соответствие между показанными на рисунке 1 видами знаний, которые необходимо представлять в базе знаний ИС, и средствами интегрированной модели, используемыми для их представления.

Конкретные знания о ПО, как и знания о моделируемом в ИС фрагменте действительности или контексте задачи представляются в семантической сети SN , включающей множество экземпляров классов (объектов) и отношений (см. рисунок 1). Для представления этого вида знаний также служат экземпляры ограничений, накладываемых на атрибуты объектов ОО-сети и хранимые в функциональной сети FN , связанной с ОО-сетью через атрибуты объектов. Для уточнения значений слотов объектов, ограничения на которые представлены в функциональной сети, используется метод недоопределённых вычислительных моделей.

Описание предметных знаний в виде правил «ЕСЛИ-ТО» средствами системы продукционных правил PS является естественным и удобным для многих областей. В интегрированной модели система продукционных правил работает над семантической сетью, состав понятий и отношений которой определяется онтологией. В связи с этим посылка и заключение любого правила могут описываться в терминах классов и отношений онтологии. Это позволяет не только удобно представлять экспертные знания в виде правил «ЕСЛИ-ТО», но и строить довольно мощные по своим изобразительным и операционным возможностям правила и с их помощью описывать на достаточно высоком уровне довольно сложные процессы вывода и обработки информации.

Знания о пространстве задач могут быть представлены в прикладной онтологии, где, во-первых, описываются типы задач и методов, предназначенных для их решения, и, во-вторых, задаётся разбиение задач на подзадачи и связь задач с методами их решения (знания второго вида также могут быть представлены средствами ОО-сети). При этом сами методы решения могут быть представлены продукционными правилами, объединёнными в группы и/или модули. Для этих целей в модель включаются средства структурирования множества правил и управления им, такие как операторы активации и деактивации групп правил, которые могут сочетаться с императивными средствами управления.

Для поддержки эффективной реализации методов и алгоритмов, а также обеспечения создания полноценных прикладных ИС в ИМПЗ-модель включены традиционные для императивной модели средства. Они применяются для разработки алгоритмов в случаях, когда их реализация другими средствами ИМПЗ-модели (например, продукционной модели) была бы

слишком неэффективной по времени и/или затратной по памяти, а также для реализации численных вычислений и обеспечения интерфейсных возможностей ИС.

Взаимосвязи интегрируемых моделей показаны на рисунке 2. Здесь видно, что сетевая и продукционная модели используют сущности, определённые в онтологической модели. В частности, структура и свойства элементов семантической сети определяются в онтологии, а посылки и заключения продукционных правил описываются в терминах классов и отношений онтологии. В свою очередь, средства оперирования, предоставляемые онтологической моделью, используют средства оперирования сетевой модели (для того, чтобы привести ОО-сеть в соответствие с текущим состоянием онтологии). Средства оперирования продукционной модели используют и функциональность сетевой модели (для сопоставления посылок правил с семантической сетью и выполнения действий над ней, определённых в заключении правил), и средства оперирования императивной модели. Кроме того, сетевая модель использует средства оперирования недоопределённой модели, а последняя – сущности, определённые в сетевой модели.

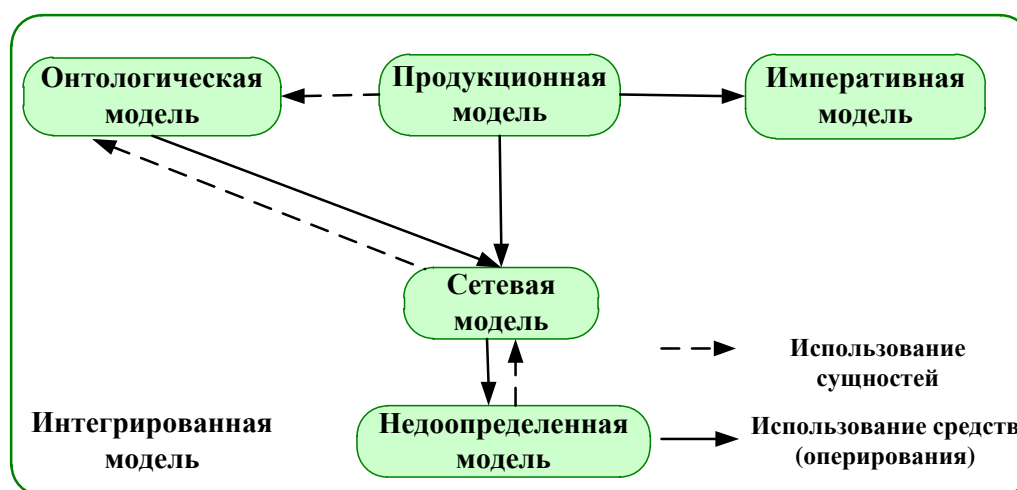


Рисунок 2 – Схема взаимодействия интегрируемых моделей

2 Язык представления и обработки знаний

Интегрированная модель представления знаний предоставляет средства для комплексного описания как декларативных, так и вычислительных свойств понятий и отношений некоторой области знаний, а также средства оперирования их экземплярами в продукционном и императивном стилях. Для доступа к этим средствам предложен язык представления и обработки знаний (ЯПОЗ). Этот язык обладает богатым набором типов данных, в том числе недоопределённых. Последние обеспечивают возможность представления недоопределённых значений атрибутов объектов в виде интервала или множества допустимых значений и оперирования ими. Благодаря этому на ЯПОЗ можно естественным образом описывать и решать задачи, в том числе, требующие сочетания логического вывода и вычислений над неточно заданными значениями.

ЯПОЗ включает средства для описания предметной области и средства для спецификации приложений, предназначенных для использования в рамках этой ПО. Первая группа средств позволяет описывать ПО в виде классов объектов и отношений онтологии, а также конструировать необходимые для этого новые типы данных и домены. Вторая группа средств предназначена для спецификации функциональности приложения, главным образом

в виде системы продукционных правил, работающей над семантической сетью, организовывать правила в группы и задавать порядок их активации. Кроме того, с целью поддержки разработки полноценных приложений в ЯПОЗ включены традиционные средства императивного программирования, а также средства, обеспечивающие разработку эргономичных пользовательских интерфейсов.

3 Инструментальные средства разработки интеллектуальных систем, основанные на интегрированной модели

Для ускорения и упрощения процесса разработки интеллектуальных систем широко применяется такой класс инструментальных систем, как программные оболочки ИС, представляющие собой интеллектуальные системы с пустой или частично заполненной базой знаний [12]. Достоинством оболочек является то, что разработка прикладных ИС на их основе не требует участия программистов, так как наполнение базы знаний входит в компетенцию инженеров знаний и экспертов.

На основе средств интегрированной модели представления знаний построены две программные оболочки – оболочка систем, основанных на знаниях (оболочка СОЗ), и оболочка порталов знаний.

3.1 Оболочка систем, основанных на знаниях

Оболочка систем, основанных на знаниях, предназначена для построения интеллектуальных систем традиционного типа, т.е. таких СОЗ, за процессы вывода и обработки информации в которых в основном отвечает система продукций. (Популярными представителями такого класса систем являются экспертные системы и системы поддержки принятия решений). В то же время благодаря тому, что такие СОЗ базируются на ИМПЗ-модели, они способны решать задачи, требующие сочетания логического вывода и вычислений над неточно заданными значениями.

Рассматриваемая оболочка включает типовую СОЗ с пустой базой знаний и средства спецификации базы знаний. В типовую СОЗ, кроме базы знаний, также входят интерфейс, обеспечивающий взаимодействие полученной системы с конечным пользователем и внешними подсистемами, и программное ядро, обеспечивающее реализацию процессов вывода и обработки информации в создаваемых СОЗ.

Базу знаний прикладной СОЗ составляет прикладная онтология, семантическая сеть, функциональная сеть, а также система продукций, определяющая основные процессы вывода и обработки информации в СОЗ.

Связанные между собой общими данными семантическая и функциональная сети играют роль основной памяти системы. В семантической сети представлены декларативные знания об объектах и отношениях между ними, в функциональной сети – вычислительные отношения, заданные на атрибутах объектов. Вид семантической и функциональной сети определяется понятиями и отношениями, введенными инженером знаний на этапе конструирования онтологии.

Программное ядро СОЗ включает интерпретатор продукционных правил, «императивную машину», отвечающую за выполнение команд обработки данных и управления, реализуемых в императивной парадигме программирования, семантический процессор, обеспечивающий выполнение всего комплекса операций над семантической сетью, и Н-процессор, реализующий процесс удовлетворения ограничений в функциональной сети.

Средства настройки базы знаний включают редактор онтологий, предоставляющий удобный интерфейс к менеджеру онтологий, отвечающему за создание, редактирование он-

тологии и доступ к ней; редактор семантической сети, обеспечивающий удобный интерфейс к семантическому процессору, с помощью которого создаётся начальное наполнение семантической сети, задающее контекст решения задачи или исходное состояние объекта (системы) управления, а также редактор правил, с помощью которого с использованием понятий и отношений, определённых в онтологии, создаются продукционные правила.

3.2 Оболочка портала знаний

Интегрированная модель представления знаний и базирующиеся на ней технологии могут быть применены не только для создания широкого класса систем, основанных на знаниях, но и для решения другой важной задачи – проблемы эффективного доступа к знаниям и информационным ресурсам, представленным в сети Интернет. Для этих целей предложена оболочка портала знаний (ПЗ), которая позволяет создавать специализированные системы – интернет-порталы, обеспечивающие интеграцию и систематизацию знаний и информационных ресурсов по требуемой тематике, а также содержательный доступ к ним.

Информационная модель такого портала базируется на интегрированной модели представления знаний, что позволяет представлять в нём разнородные знания и данные. Для удобства разработки в информационной модели выделяются два уровня. Первый уровень составляет онтология портала знаний, задающая понятия и существующие между ними отношения моделируемой области знаний, включая типологию соответствующих ей информационных ресурсов. Второй уровень представляет собой объектно-ориентированную семантическую сеть, содержащую информационные объекты (экземпляры понятий онтологии), которые связаны отношениями, определёнными в онтологии, т.е. второй уровень выступает как информационное наполнение или контент портала знаний. На основе данной информационной модели организуются удобная навигация по знаниям и информационным ресурсам, интегрированным в портале знаний, и содержательный поиск требующихся информационных объектов.

Оболочка портала знаний строится по тем же принципам, что и оболочка СОЗ. Она представляет собой типовой ПЗ и включает пользовательский интерфейс, базу знаний, программное ядро, состоящее из менеджера онтологий, семантического процессора и Н-процессора, а также средства настройки БЗ.

Управляемый онтологией пользовательский интерфейс представляет собой веб-приложение, обеспечивающее навигацию по знаниям и информационным ресурсам, представленным в онтологии и контенте ПЗ, а также содержательный поиск требующейся информации. Во время своей работы пользовательский интерфейс обращается к менеджеру онтологий и семантическому процессору, «поставляющим» ему необходимые данные.

Средства настройки базы знаний включают редакторы онтологий и контента, а также коллекционер онтологической информации (коллекционер ОИ).

Редактор контента, работающий под управлением онтологии, предоставляет экспертам удобные и разнообразные средства для управления контентом портала знаний.

Коллекционер ОИ предназначен для повышения степени автоматизации разработки контента ПЗ. Он выполняет сбор онтологической информации (расширенных метаданных) об информационных ресурсах и публикациях по тематике портала и согласованное пополнение ими контента ПЗ.

Вся работа по созданию конкретного портала знаний состоит в задании необходимых предметных знаний, т.е. онтологии области знаний и соответствующих ей информационных объектов и связей между ними, составляющих контент ПЗ, а также настройке пользовательского веб-интерфейса.

Так как указанные выше редакторы и коллекционер ОИ непосредственно входят в состав портала знаний, благодаря этому они могут использоваться не только на этапе его создания, но и во время эксплуатации, обеспечивая поддержку его системы знаний в актуальном состоянии.

4 Технология разработки интеллектуальных систем, ориентированная на экспертов

Рассмотренные выше оболочки СОЗ и порталов знаний являются эффективными средствами построения интеллектуальных систем указанных типов. Однако есть ещё одна нерешённая проблема, которая состоит в том, что оболочка ИС, даже основанная на описанной выше интегрированной модели представления знаний, является всё ещё трудной для непосредственного использования экспертом. Это объясняется тем, что она содержит только универсальные средства представления знаний, которые далеки от области знаний эксперта и поэтому требуют участия в разработке конкретных систем инженеров знаний.

Представляется очевидным, что чем меньше будет разрыв между знаниями и представлениями эксперта и средствами представления знаний инструментальной системы, тем легче ими будет пользоваться эксперту. Поэтому специализация оболочки ИС на определённые классы областей знаний и задач, т.е. включение в её состав «знаниевых» компонентов, содержащих базовый набор понятий, относящихся к заданному классу областей знаний, а также программных компонентов, поддерживающих требуемую в ИС функциональность, сделало бы её более понятной экспертам и пригодной для использования ими без привлечения инженеров знаний.

На описанной выше идее специализированных оболочек строится концепция технологии разработки СОЗ и порталов знаний, ориентированная на экспертов. В соответствии с этой концепцией общая архитектура инструментальной поддержки разработки интеллектуальных систем выглядит так, как показано на рисунке 3.

Базис инструментальной поддержки составляет базирующаяся на ИМПЗ-модели и языке ЯПОЗ интегрированная система представления и обработки знаний, которая представляет собой набор языковых и программных средств, реализующих интегрированную модель представления знаний.

Следующий уровень включает универсальные оболочки, предназначенные для построения СОЗ и порталов знаний. Оболочки этого уровня предоставляют универсальные средства для построения систем данного типа и предназначены для инженеров знаний.

Самый верхний уровень образуют специализированные оболочки, ориентированные на определённые области знаний и классы задач, и благодаря этому пригодные для использования экспертами. Причем специализированные оболочки строятся инженерами знаний с помощью средств разработки БЗ, предоставляемых универсальными оболочками.

В соответствии с изложенной концепцией специализированная оболочка СОЗ включает базовые онтологии, на основе которых строится онтология области знаний СОЗ и онтология задач, а также набор специализированных продукционных модулей и библиотек процедур и функций (на языке ЯПОЗ), реализующих решение базовых задач.

Создание конкретной СОЗ на основе такой оболочки заключается в построении онтологии области СОЗ путём расширения и развития базовых онтологий, построении дополнительных продукционных модулей, реализующих задачи, специфичные для разрабатываемой СОЗ, а также наполнении декларативной части базы знаний необходимым набором экземпляров объектов и отношений, представляющих универсальные факты или начальное состояние объекта управления.



Рисунок 3 – Общая архитектура инструментальной поддержки разработки интеллектуальных систем

В качестве примера специализированной оболочки СОЗ приведём оболочку СОЗ медицинской диагностики (рисунок 4) [13]. Эта оболочка включает онтологию медицинской диагностики и лечения, которая содержит понятия и отношения, общие для всех диагностических медицинских систем.

В рамках данного подхода была создана ещё одна специализированная оболочка – оболочка СОЗ технической диагностики технологической инфраструктуры предприятия [14]. Эта оболочка содержит верхний уровень онтологии технологической инфраструктуры предприятия, а также библиотеку процедур и функций для анализа временных рядов, которые могут использоваться в правилах, описывающих мониторинг и диагностику технологической инфраструктуры предприятия.

Аналогичным образом строятся специализированные оболочки порталов знаний. Каждая такая оболочка представляет собой типовой ПЗ с набором базовых онтологий, описывающих основные сущности некоторого класса областей знаний, и пустым контентом.



Рисунок 4 – Специализированная оболочка СОЗ медицинской диагностики

Разработка конкретного портала знаний с помощью специализированной оболочки заключается, главным образом, в настройке его на требуемую область знаний, которая состоит в построении онтологии области знаний ПЗ на основе включённых в оболочку базовых онтологий и заполнении контента под управлением построенной онтологии. Эту работу уже могут выполнять сами эксперты без привлечения инженеров знаний.

Как было отмечено выше, специализированные оболочки строятся с помощью средств, предоставляемых универсальными оболочками. В частности, средствами оболочки портала знаний были построены две оболочки – оболочка для разработки порталов научных знаний (ПНЗ-оболочка) [15] и оболочка для создания многоязычных тезаурусов (Т-оболочка) [16].

Эти оболочки отличаются, главным образом, набором базовых онтологий. Так, в ПНЗ-оболочку (рисунок 5) включены две базовые онтологии – онтология научного знания, содержащая базовые понятия, задающие структуры для описания области знаний портала, и онтология научной деятельности, содержащая базовые понятия, служащие для описания различных аспектов научно-исследовательской деятельности. В то время как Т-оболочка включает онтологию представления тезауруса, основанную на стандартах построения многоязычных информационно-поисковых тезаурусов. Кроме того, в Т-оболочку за ненадобностью не включён коллекционер онтологической информации.

Специализированные оболочки могут быть отправной точкой для построения ещё более специализированных оболочек. Например, на основе ПНЗ-оболочки можно построить оболочку для разработки порталов знаний, предназначенных для поддержки инновационной деятельности, расширив набор базовых онтологий ПНЗ-оболочки онтологией производственной и инновационной деятельности.

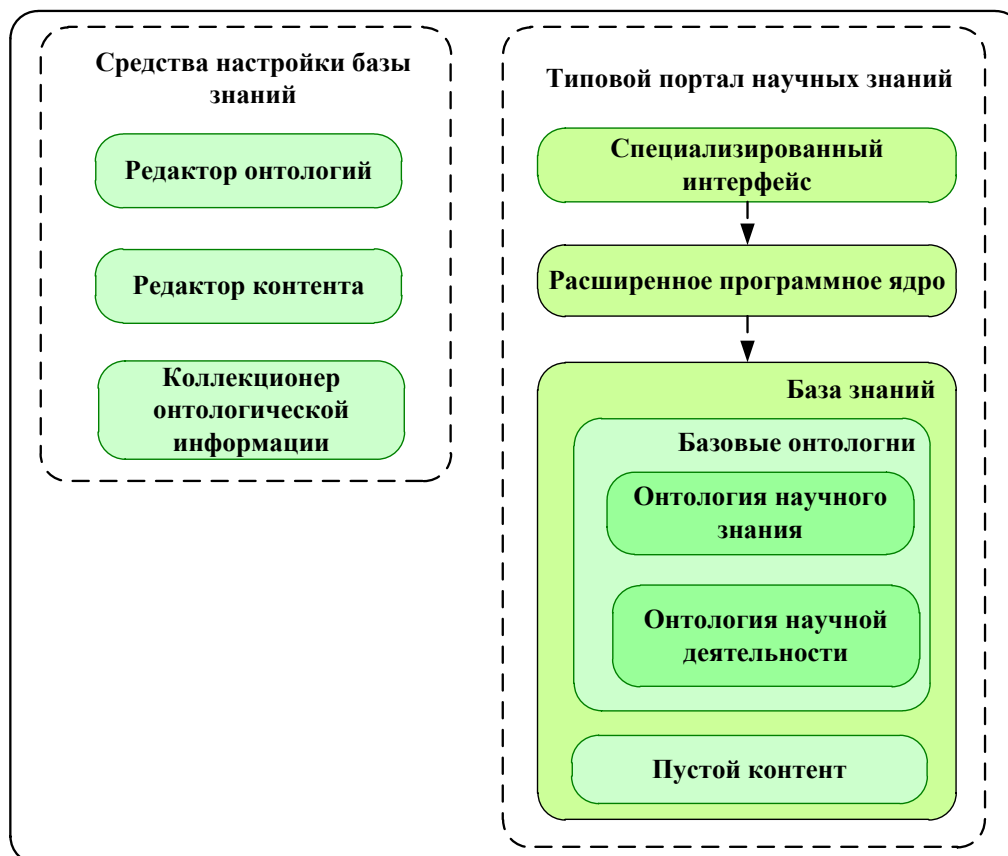


Рисунок 5 – Оболочка портала научных знаний

Заключение

В статье представлена технология построения интеллектуальных систем, ориентированная на экспертов. В качестве базиса представления знаний в ИС предлагается модель, интегрирующая взаимодополняющие друг друга методы и средства представления и обработки знаний следующих моделей: онтологической, сетевой, продукционной, недоопределённой и императивной. Набор языковых и программных средств, реализующих интегрированную модель, служит базисом инструментальных средств, которые могут применяться для разработки широкого класса информационных и интеллектуальных систем. Для того, чтобы сделать эти средства доступными для непосредственного использования экспертами, была предложена технология построения интеллектуальных систем, базирующаяся на специализированных оболочках, настроенных на определённые области знаний и классы задач. При этом сами специализированные оболочки могут создаваться инженерами знаний путем частичного заполнения баз знаний соответствующих универсальных оболочек с использованием имеющихся в них средств описания предметных знаний.

Продуктивность данного подхода подтверждается тем, что в его рамках были созданы специализированные оболочки, которые успешно применялись при создании практически полезных ИС. В частности, оболочка СОЗ технической диагностики технологической инфраструктуры предприятия применялась при создании системы поддержки принятия решений на нефтегазодобывающем предприятии [17], с помощью ПНЗ-оболочки были созданы порталы знаний по компьютерной лингвистике и археологии [18], а с помощью Т-оболочки – электронный двуязычный тезаурус по компьютерной лингвистике [19].

Благодарности

Автор благодарен коллегам из ИСИ СО РАН, принимавшим участие в разработке и реализации инструментальных средств, поддерживающих представленную в статье технологию, а также всем, кто участвовал в апробации данной технологии при построении конкретных приложений. Автор также выражает благодарность РФФИ за финансовую поддержку данной работы (проект № 13-07-00422).

Список источников

- [1] *Гаврилова, Т.А.* Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2001. – 384 с.
- [2] *Осипов, Г.С.* Построение моделей предметных областей. Неоднородные семантические сети / Г.С. Осипов // Известия АН СССР. Техническая кибернетика. –1990. – №5. – С. 32–45.
- [3] *Минский, М.* Структура для представления знания / М. Минский // Психология машинного зрения. – М.: Мир, 1978. – С. 249-338.
- [4] *Поспелов, Д.А.* Продукционные модели / Д. А. Поспелов // Искусственный интеллект. В 3-х кн. Кн.2. Модели и методы: Справочник / Под. Ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – С. 49-56.
- [5] *Ушаков, Д.М.* Системы программирования в ограничениях (обзор) / Д.М. Ушаков, В.В. Телерман // Системная информатика: Сб. науч. тр. Под. Ред. И.В. Поттосина. Вып.7: Проблемы теории и методологии создания параллельных и распределенных систем. – Новосибирск: Наука, 2000. –С. 275–310.
- [6] *Нариньяни, А.С.* Недоопределённость в системах представления и обработки знаний / А.С. Нариньяни // Изв. АН СССР. Техн. кибернетика. – 1986. – №5. – С. 3-28.
- [7] *Нариньяни, А.С.* Программирование в ограничениях и недоопределённые модели / А.С. Нариньяни, В.В. Телерман, Д.М. Ушаков, И.Е. Швецов // Информационные технологии, 1998. – №7. – С. 13-22.
- [8] *Guarino, N.* Formal Ontology in Information Systems / N. Guarino // Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6–8 June 1998. – Amsterdam: IOS Press, 1998. – P. 3–15.
- [9] *Боргест, Н.М.* Онтологии: современное состояние, краткий обзор / Н.М. Боргест, М.Д. Коровин // Онтология проектирования. – 2013. – №2(8). – С. 49-55.
- [10] *Копайгородский, А.Н.* Применение онтологий в семантических информационных системах / А.Н. Копайгородский. // Онтология проектирования. – 2014. – №4(14). – С.78-89.
- [11] *Baader, F.* The Description Logic Handbook: Theory, Implementation, Applications / F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. – Cambridge, 2003. – 574 p.
- [12] *Артемьева, И.Л.* Специализированные оболочки интеллектуальных систем для сложно-структурированных предметных областей / И.Л. Артемьева // Труды 11-й национальной конференции по искусственному интеллекту с международным участием – КИИ-2008. – М: ЛЕНАНД, 2008. -Т.1. - С. 95-103.
- [13] *Загорулько, Г.Б.* Экспертная система поддержки диагностики, профилактики и лечения элементозов на основе коррекции питания / Г.Б. Загорулько, А.М. Гончар, М.Н. Рубан, А.Н. Рябков // Труды 10-й национальной конференции по искусственному интеллекту с международным участием – КИИ-2006. – М.: Физматлит, 2006. – Т.3. – С. 849-857.
- [14] *Zagorulko, Y.* Architecture of Extensible Tools for Development of Intelligent Decision Support Systems / Y. Zagorulko, G. Zagorulko // New Trends in Software Methodologies, Tools and Techniques. Proceedings of the Tenth SoMeT_11. Eds.: Hamido Fujita and Tatiana Gavriloa. – IOS Press, Amsterdam, 2011. – P. 253-263.
- [15] *Загорулько, Ю.А.* Технология разработки порталов научных знаний / Ю.А. Загорулько // Программные продукты и системы. – 2009. – №4. – С. 25-29.
- [16] *Загорулько, Ю.А.* Программная оболочка для построения многоязычных тезаурусов предметных областей, ориентированная на экспертов / Ю.А. Загорулько, О.И. Боровикова // Труды 13-й национальной конференции по искусственному интеллекту с международным участием – КИИ-2012. – Белгород: Изд-во БГТУ, 2012. -Т.4. - С. 76-83.
- [17] *Загорулько, Ю.А.* Онтологический подход к разработке системы поддержки принятия решений на нефтегазодобывающем предприятии / Ю.А. Загорулько, Г.Б. Загорулько // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2012. Т.10. Вып. 1. – С. 121-128.
- [18] *Ануреев, И.С.* Модели и методы построения информационных систем, основанных на формальных, логических и лингвистических подходах / И.С. Ануреев, Т.В. Батура, О.И. Боровикова, Ю.А. Загорулько, И.С. Кононенко, А.Г. Марчук, П.А. Марчук, Ф.А. Мурзин, Е.А. Сидорова, Н.В. Шилов // Отв. ред. А.Г. Марчук. – Новосибирск: Изд-во СО РАН, 2009. – 330 с.

- [19] *Загорулько, Ю.А.* Методологические аспекты разработки электронного русско-английского тезауруса по компьютерной лингвистике / Ю.А. Загорулько, О.И. Боровикова, И.С. Кононенко, Е.Г. Соколова // Информатика и ее применения. – 2012. – Т. 6. – №3. – С. 22–31.

SEMANTIC TECHNOLOGY FOR DEVELOPMENT OF INTELLIGENT SYSTEMS ORIENTED ON EXPERTS IN SUBJECT DOMAIN

Yu.A. Zagorulko

*A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
zagor@iis.nsk.su*

Abstract

The semantic models and technologies for building intelligent systems oriented on experts in a subject domain are considered in the paper. The integrated model combining various mutually complementary methods and means of knowledge representation and processing of such models as ontological, network, production, subdefinite and imperative is developed and suggested as basis for knowledge representation in intelligent systems. Suite of language and programming tools built on the basis of this model can be used for development of wide array of such kind of systems. In order to make these tools accessible to use immediately by experts, the technology for development of intelligent systems is suggested. This technology is based on specialized program shells, which are adjusted to certain knowledge area and class of tasks. These shells are built by knowledge engineers who perform a partial filling of knowledge bases of corresponding universal shells using the knowledge base of development tools of these shells.

Key words: *integrated knowledge representation model, ontology, intelligent system, knowledge portal, specialized program shell.*

Acknowledgment

The author would like to acknowledge the support of his colleagues from IIS SB RAS, who participated in design and realization of software tools supporting the technology presented in the paper, as well as everyone who participated in the approbation of the said technology in building endpoint applications. The author is also grateful for the financial support of RFBR for this work (Project No. 13-07-00422).

References

- [1] *Gavrilova, T.A.* Bazy znaniy intellektual'nyh system [Databases of the intellectual systems] / T.A. Gavrilova, V.F. Horoshevskij. – SPb.: Piter, 2001. – 384 p. (In Russian).
- [2] *Osipov, G.S.* Postroenie modelej predmetnyh oblastej. Neodnorodnye semanticheskie seti [The construction of domain models. Inhomogeneous semantic nets] / G.S. Osipov // Izvestija AN SSSR. Tehnicheskaja kibernetika. – 1990. – №5. – pp. 32–45. (In Russian).
- [3] *Minskij, M.* Struktura dlja predstavlenija znanija [The structure for the representation of knowledge] / M. Minskij // Psihologija mashinnogo zrenija. – M.: Mir, 1978. – P. 249-338. (In Russian).
- [4] *Pospelov, D.A.* Produkcionnye modeli [Productional models] / D.A. Pospelov // Iskusstvennyj intellekt. V 3-h kn. Kn.2. Modeli i metody: Spravochnik / Pod. Red. D.A. Pospelova. – M.: Radio i svjaz', 1990. – pp. 49-56. (In Russian).
- [5] *Ushakov, D.M.* Sistemy programmirovanija v ogranichenijah (obzor) [Systems of constraint programming (review)] / D.M. Ushakov, V.V. Telerman // Si-stemnaja informatika: Sb. nauch. tr. Pod. Red. I.V. Pottosina. Vyp.7: Problemy teorii i metodologii so-zdanija parallel'nyh i raspredeleennyh sistem. – Novosibirsk: Nauka, 2000. – pp. 275–310. (In Russian).
- [6] *Narin'jani, A.S.* Nedoopredelennost' v sistemah predstavlenija i obrabotki znaniy [Underdetermined systems knowledge representation and processing] / A.S. Narin'jani // Izv. AN SSSR. Tehn. kibernetika. – 1986. – №5. – pp. 3-28. (In Russian).
- [7] *Narin'jani, A.S.* Programmirovanie v ogranichenijah i nedoopredeljonnye modeli [Constraint programming and underdetermined model] / A.S. Narin'jani, V.V. Telerman, D.M. Ushakov, I.E. Shvecov // Informacionnye tehnologii, 1998. – №7. – pp. 13-22. (In Russian).

- [8] **Guarino, N.** Formal Ontology in Information Systems / N. Guarino // Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6–8 June 1998. – Amsterdam: IOS Press, 1998. – P. 3–15.
- [9] **Borgest, N.M.** Ontologii: sovremennoe sostojanie, kratkij obzor [Ontologies: current state and brief review] / N.M. Borgest, M.D. Korovin // Ontologia proektirovaniya. – 2013. – No. 2(8). – pp. 49-55. (In Russian).
- [10] **Kopajgorodskij, A.N.** Primenenie ontologij v semanticheskikh informacionnyh sistemah [The use of ontologies in semantic information systems] / A.N. Kopajgorodskij. // Ontologija proektirovaniya. – 2014. – No. 4(14). – pp.78-89. (In Russian).
- [11] **Baader, F.** The Description Logic Handbook: Theory, Implementation, Applications / F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider. – Cambridge, 2003. – 574 p.
- [12] **Artem'eva, I.L.** Specializirovannye obolochki intellektual'nyh sistem dlja slozhno-strukturirovannyh predmetnyh oblastej [Specialized shells of intellectual systems for complex-structured domains] / I.L. Artem'eva // Trudy 11-j nacional'noj konferencii po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2008. – M.: LENAND, 2008. -Vol.1. - pp. 95-103. (In Russian).
- [13] **Zagorulko, G.B.** Jekspertnaja sistema podderzhki diagnostiki, profilaktiki i lechenija jelementozov na osnove korekcii pitaniya [Expert system support the diagnosis, prevention and treatment of elementosis based on the correction power] / G.B. Zagorulko, A.M. Gonchar, M.N. Ruban, A.N. Rjabkov // Trudy 10-j nacional'noj konferencii po iskusstvennomu intellektu s mezhdunarodnym uchastiem – KII-2006. – M.: Fizmatlit, 2006. -T.3. - p. 849-857. (In Russian).
- [14] **Zagorulko, Yu.** Architecture of Extensible Tools for Development of Intelligent Decision Support Systems / Yu. Zagorulko, G. Zagorulko // New Trends in Software Methodologies, Tools and Techniques. Proceedings of the Tenth SoMeT_11. . Eds.: Hamido Fujita and Tatiana Gavrilova. –I OS Press, Amsterdam, 2011. – pp. 253-263.
- [15] **Zagorulko, Yu.A.** Tehnologija razrabotki portalov nauchnyh znaniy [Technology portal development of scientific knowledge] / Ju.A. Zagorul'ko // Programmnye produkty i sistemy. – 2009. – No.4. – pp. 25-29. (In Russian).
- [16] **Zagorulko, Yu.A.** Programmnaja obolochka dlja postroenija mnogojazychnyh tezaurusov predmetnyh oblastej, orientirovannaja na jekspertov [Shell program for the construction of multilingual thesauri domains oriented experts] / Yu.A. Zagorul'ko, O.I. Borovikova // Trudy 13-j nacional'noj konferencii po iskusstvennomu intellektu s mezhdunarodnym uchastiem KII-2012. – Belgorod: Izd-vo BGTU, 2012. -Vol.4. - pp. 76-83. (In Russian).
- [17] **Zagorulko, Yu.A.** Ontologicheskij podhod k razrabotke sistemy podderzhki prinjatija reshenij na neftegazodobyvajushhem predpriyatii [An ontological approach to the development of a decision support system for oil and gas companies] / Yu.A. Zagorulko, G.B. Zagorul'ko // Vestnik Novosibirskogo gosudarstvennogo universiteta. Serija: Informacionnye tehnologii. 2012. Vol. 10. Vyp. 1. – pp. 121-128. (In Russian).
- [18] **Anureev, I.S.** Modeli i metody postroenija informacionnyh sistem, osnovannyh na formal'nyh, logiche-skih i lingvisticheskikh podhodah [Models and methods of construction of information systems based on formal logic and linguistic approaches] / I.S. Anureev, T.V. Batura, O.I. Borovikova, Yu.A. Zagorulko, I.S. Kononenko, A.G. Marchuk, P.A. Marchuk, F.A. Murzin, E.A. Sidorova, N.V. Shilov // Otv. red. A.G. Marchuk. – Novosibirsk: Izd-vo SO RAN, 2009. – 330 p. (In Russian).
- [19] **Zagorulko, Yu.A.** Metodologicheskie aspekty razrabotki jelektronnogo rusско-anglijskogo tezaurusa po komp'juternoj lingvistike [Methodological aspects of the development of e-Russian-English thesaurus on Computational Linguistics] / Yu.A. Zagorulko, O.I. Borovikova, I.S. Kononenko, E.G. Sokolova // Informatika i ejo primenenija. – 2012. – Vol. 6. – No. 3. – pp. 22–31. (In Russian).

Сведения об авторе



Загорулько Юрий Алексеевич, 1957 г. рождения. Окончил Новочеркасский политехнический институт им. С. Орджоникидзе в 1979 г., к.т.н. (1989). Заведующий лабораторией Института систем информатики им. А.П. Ершова СО РАН, доцент кафедры программирования и систем информатики Новосибирского государственного университета. Член Российской ассоциации искусственного интеллекта. В списке научных трудов более 200 публикаций в области искусственного интеллекта, разработки интеллектуальных систем, инженерии знаний, онтологического моделирования и компьютерной лингвистики.

Zagorulko Yuriy Alekseevich (b. 1957) graduated from the Novocherkassk Polytechnic Institute in 1979, PhD (1989). He is the Head of laboratory at A.P. Ershov Institute of Informatics Systems, Siberian Branch of the Russian Academy of Science (Novosibirsk city), associate professor at Novosibirsk State University (Department of Mechanics and Mathematics, Department of Information Technologies), He is member of Russian Association of Artificial Intelligence, author of more than 200 publications in the field of AI, knowledge and ontology engineering, intelligent system development and computational linguistics.