

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin
(editors)

**Information Models
of
Knowledge**

**ITHEA[®]
KIEV – SOFIA
2010**

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)

Information Models of Knowledge

ITHEA®

Kiev, Ukraine – Sofia, Bulgaria, 2010

ISBN 978-954-16-0048-1

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

ITHEA IBS ISC: 19.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligence metasynthesis and knowledge processing in intelligent systems; Formalisms and methods of knowledge representation; Connectionism and neural nets; System analysis and synthesis; Modelling of the complex artificial systems; Image Processing and Computer Vision; Computer virtual reality; Virtual laboratories for computer-aided design; Decision support systems; Information models of knowledge of and for education; Open social info-educational platforms; Web-based educational information systems; Semantic Web Technologies; Mathematical foundations for information modeling of knowledge; Discrete mathematics; Mathematical methods for research of complex systems.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Ukraine

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co., Bulgaria

ISBN 978-954-16-0048-1

C/o Jusautor, Sofia, 2010

КОМПЬЮТЕРНЫЕ ОНТОЛОГИИ И ОНТОЛОГО-УПРАВЛЯЕМАЯ АРХИТЕКТУРА ИНФОРМАЦИОННЫХ СИСТЕМ

Николай Петренко

Abstract: В работе описана архитектура двух основных компонент онтологической системы с обработкой естественно-языковых объектов – компьютерной онтологии и программно-аппаратных средств с онтолого-управляемой архитектурой.

Keywords: компьютерные онтологии, онтолого-управляемая архитектура.

ACM Classification Keywords: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods.

Введение

Системный подход к познанию ориентирует аналитика на рассмотрение любой ПДО с позиций закономерностей системного целого и взаимодействия составляющих его частей. Системность знаний исходит из многоуровневой иерархической организации любой сущности, то есть все объекты, процессы и явления можно рассматривать как множество более мелких подмножеств (признаков, деталей) и, наоборот, любые объекты можно (и нужно) рассматривать как элементы более высоких классов обобщений [Палагин, 2009, а].

Согласно [Аверкин, 1992] под *анализом* понимается вид исследования, при котором реальный или мыслимый объект расчленяется на составляющие части (элементы) и исследуются эти элементы и связи между ними. Анализ предметной области представляет особый вид научной деятельности, в результате которой строится интерпретационная модель предметных знаний (в широком смысле). В процессе анализа последние делятся на инвариантные и прагматические знания, концептуальные составляющие которых представляют онтологические знания ПДО.

Компьютерные онтологии

Начало 90-х годов прошлого столетия считается зарождением *парадигмы компьютерных онтологий*. Она была сформулирована как попытка сгладить (и по возможности устранить) всё чаще проявляющиеся разного рода противоречия при функционировании и внедрении интеллектуальных систем с использованием баз знаний предметных областей. Ярким представителем таких систем на то время были экспертные системы (ЭС). Они успешно и эффективно функционировали в пределах одного коллектива и на уровне коммерческих образцов. У разработчиков постоянно возникал вопрос: “Как обеспечить продвижение этих образцов к конечному пользователю?”. Были предложены разработки оболочек – “пустых” ЭС и ряд других новшеств. Но они не имели решающего значения. Напрашивался вывод, что для эффективного функционирования ЭС на самом важном этапе “жизненного цикла” – функционирования у конечного пользователя при решении реальных задач – необходимо к каждой ЭС “приложить” эксперта в соответствующей ПДО. Необходимость присутствия эксперта объяснялась, в частности, быстротечностью изменения знаний во многих предметных областях и соответственно необходимостью обновления базы знаний ЭС в “реальном времени”. В тех же ПДО, где знания обладали относительной инвариантностью, ЭС продолжали эффективно функционировать.

Следующим этапом развития парадигмы компьютерных онтологий являлось задекларированное в [Guarino, 1998] построение ИС с онтолого-управляемой архитектурой (или онтолого-управляемых ИС).

Сказанное выше и ряд других факторов “подтолкнули” к разработке парадигмы компьютерных онтологий, основные принципы которой были сформулированы в [Gruber, 1993].

1. Доходчивость, ясность (Clarity). Термины (и понятия) онтологии должны отражать реальную действительность. Их символные обозначения (знаки) должны формироваться на основе

- общепринятых правил в семиотике и должны выражать общепринятые смыслы реальных объектов. В свою очередь, эти смыслы извлекаются из общепринятых определений терминов (понятий), зафиксированных в толковых словарях, различных глоссариях ПдО. Суждения, входящие в определения, формализуются на основе формального общепринятого аппарата в виде тождественно истинных логических аксиом.
2. Обоснованность, связность (Coherency). Формирование начального набора понятий онтологии и их добавление должно быть обоснованным, определяемым, в первую очередь, требованиями предполагаемой совокупности решаемых задач. Логические аксиомы начального набора понятий должны быть непротиворечивыми. Для этого должен быть предусмотрен механизм логического вывода, который, в том числе, проверяет на непротиворечивость добавляемые аксиомы и выводимые в онтологии утверждения.
 3. Расширяемость (Extendibility). Ядром онтологии являются первоначально введенные (спроектированные) понятия и описывающие их аксиомы. В онтологии должен быть предусмотрен механизм расширения (ограничения) совместно используемых словарей понятий без нарушения целостности системы.
 4. Минимальное влияние кодирования (Minimal encoding bias). В онтологической системе (ОнС) должен быть реализован принцип совместного использования онтологий, который предполагает: спецификацию онтологии на уровне представления, а не символического кодирования; запись такой спецификации на общепринятом и платформонезависимом языке описания онтологий можно передать для использования любому программному агенту.
 5. Минимальные онтологические обязательства (Minimal ontological commitment). Этот принцип перекликается с принципами обоснованности и расширяемости/ограничения. Важно, чтобы множество понятий онтологии отображало концептуальную структуру ПдО, относительно стабильную на протяжении “жизненного цикла” ОнС. А последняя предоставляла возможность расширения или специализации отдельных ветвей онтологического графа (онтографа). Отделение концептуальных знаний от знаний, выраженных фактами, является стратегией построения ОнС, а точнее – онтологических баз знаний.

В отличие от онтологии как философской дисциплины остановимся на рассмотрении компьютерной (формальной) онтологии предметной области и онтологии домена предметных областей. Последняя (в том числе) имеет важное значение для объединения (интеграции) концептуальных знаний близких предметных областей или реализации технологии системной интеграции трансдисциплинарных научных знаний. Мы также выделяем *начальную онтологию ПдО*, являющуюся инициализирующей доминантой при реализации технологии автоматизированного построения онтологии ПдО.

Известны общепринятые определения онтологии ПдО Грубера, Гуарино, Гавриловой, на основе которых разные исследователи формулируют свои, частные определения онтологии ПдО – в соответствии с их конкретной областью профессиональных интересов¹. Тематика и практическая направленность наших исследований (онтология как средство построения баз знаний и трансдисциплинарных научных исследований) предопределяет следующее определение компьютерной онтологии ПдО.

Компьютерная онтология ПдО – это:

- иерархическая структура конечного множества понятий, описывающих заданную предметную область;
- структура представляет собой онтограф, вершинами которого являются понятия, а рёбрами – семантические отношения между ними;
- понятия и отношения интерпретируются в соответствии с общезначимыми функциями интерпретации, взятыми из электронных источников знаний из заданной ПдО;

¹ Таких определений известно несколько десятков.

- дополнительные интерпретации понятий и отношений определяются аксиомами и ограничениями их области действия;
- формально онтограф описывается на одном из языков описания онтологий;
- функции интерпретации и аксиомы описаны в некоторой подходящей формальной теории.

Общепринято онтологию некоторой ПДО формально представлять упорядоченной тройкой [Палагин, 2005; Гаврилова, 2001; Guarino, 1998]

$$O = \langle X, R, F \rangle, \quad (1)$$

где X, R, F – конечные множества соответственно: X – концептов (понятий, терминов) предметной области, R – отношений между ними, F – функций интерпретации X и/или R .

Рассмотрение граничных случаев множеств (1): $R = \emptyset$; $R \neq \emptyset$; $F = \emptyset$; $F \neq \emptyset$ во всех четырёх комбинациях значений R и F даёт различные варианты онтологических конструкций, начиная от простого словаря и таксономии до формальной структуры концептуальной базы знаний для высокоинтеллектуальных знание-ориентированных систем. По своей функциональной полноте и степени формальности различаются три вида онтологий: простая, полная (или строгая) и множество промежуточных или неполных онтологий.

Простая онтология – это такая онтология, в которой $R = \emptyset$; $F = \emptyset$. Она служит (в основном) для однозначного восприятия научным сообществом понятий в соответствующей прикладной области.

Строгая или *полная онтология* ($R \neq \emptyset$; $F \neq \emptyset$) – это такая онтология, в которой множества концептов и концептуальных отношений максимально полные, а к функциям интерпретации добавляются аксиомы, определения и ограничения. При этом описания всех компонент представлены на некотором формальном языке, доступном для их интерпретации компьютером. Схема формальной модели полной онтологии описывается четвёркой:

$$O = \langle X, R, F, A(D, Rs) \rangle, \quad (2)$$

где X – множество концептов;

R – множество концептуальных отношений между ними;

$F : X \times R$ – конечное множество функций интерпретации, заданных на концептах и/или отношениях;

A – конечное множество аксиом, которые используются для записи всегда истинных высказываний (определений и ограничений);

D – множество дополнительных определений понятий;

Rs – множество ограничений, определяющих область действия понятийных структур.

Полная (или компьютерная) онтология является (формальным) выражением концептуальных знаний о предметной области и по своей значимости сопоставима с базой знаний знание-ориентированной информационной системы (ЗОИС), а её построение является специфической формой человеческого творчества. Творческий процесс можно представить совокупностью операций-процедур с суждениями, утверждениями, понятиями и отношениями между ними. А последние являются фундаментом, основой для построения составной части научной теории – онтологической базы знаний в заданной предметной области. При этом такие знания описываются в декларативной форме.

Известно, что проектируемые средства информатики в соответствии с их проблемной ориентацией базируются на некоторой совокупности фундаментальных принципов, методик и алгоритмов. Также известны многочисленные разработки в области обработки ЕЯТ [Широков, 2005 и др.], в которой ввиду огромной сложности решаемых задач принимаются те или иные ограничения, снижающие эффективность средств обработки. Автором были рассмотрены и проанализированы разработки, в которых формальная

онтология играет существенную роль. В результате были выявлены главные недостатки, снижающие эффективность реализации отдельных процедур и конечных результатов в целом.

Известны общие совокупности методов, принципов, процедур и атрибутов системного анализа как научного познания в любой ПдО [Згуровский, 2005 и др.]. При конкретизации ПдО конкретизируются и средства системного анализа. Парадигма компьютерных онтологий, развиваемая во взаимодействии со средствами и методами системного анализа, положила начало развитию новой ветви средств и методов системного анализа ПдО – системно-онтологического анализа (или подхода).

Центральной идеей системно-онтологического подхода (СОП) является разработка онтологических средств поддержки решения прикладных задач – полифункциональной онтологической системы. Такая система (точнее, её концептуальная часть) описывается двойкой (3) и представляет онтологию ПдО, состоящую из онтологии объектов и онтологии процессов, и онтологию задач [Палагин, 2009, а].

$$O_{HC} = \langle O^{поо}(O^O, O^П), O^З \rangle \quad (3)$$

На рис. 1 представлена схема онтологий-компонентов предметной области и проблемного пространства (ПрП). ПрП – это модель всех таких аспектов или компонент ПдО, с которыми связаны (опосредственно или непосредственно) знания, требующиеся при решении различных задач в этой ПдО. Всякое ПрП состоит из двух блоков: инвариантной (относительно неизменной) части и множества изменяемых частей, соответствующих отдельным задачам. В составе инвариантной части, например в методологии SMEE (Structured Methodology for Elicitation of Expertise), выделяют семь типов компонент: объекты, инструменты, операторы, операции, конечные продукты, побочные продукты и ограничения [Методический комплекс]. Эти типы компонент – суть понятия, которые хорошо группируются в онтологии объектов и процессов, представленных на рис. 1. На нём приняты следующие обозначения:

- O^O – онтология множества объектов (понятий, концептов) ПдО, которая рассматривается как иерархическая структура классов, подклассов и элементов классов;
- $O^П$ – онтология множества процессов ПдО, которая рассматривается как иерархическая структура процессов, подпроцессов, действий и операций;
- $O^З$ – онтология совокупности задач (типовых наборов), которые могут быть поставлены и решены в ПдО. Рассматривается как иерархическая структура задач, подзадач, процедур и операторов.

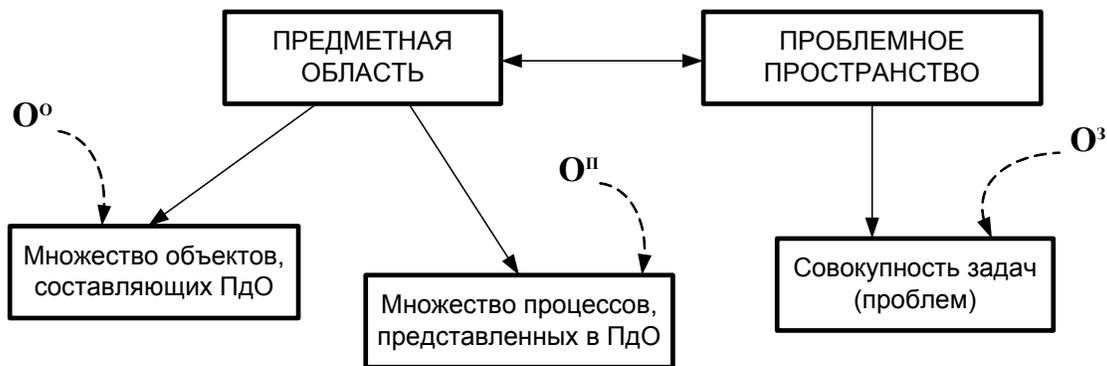


Рис. 1. Схема онтологий-компонентов предметной области

Функционально полная (с практической точки зрения) компьютерная онтология ПдО содержит декларативные (O^O и $O^П$) и процедурные ($O^З$) знания. Эти знания, взаимодействуя между собой, обеспечивают знание-ориентированную поддержку процесса решения типового набора задач пользователя. Инициальными являются процедурные знания, а точнее – компоненты онтологии задач – обобщенный алгоритм решения типового набора задач (представленный в онтологическом виде), формирователь методов решения типового набора задач и блок решателя задач. После того, как выбран метод и адаптирован алгоритм для решения целевой задачи, блок решателя задач формирует исходные

данные и последовательно запускает соответствующие процессы из онтологии процессов, которая, в свою очередь, обращается к соответствующим концептам онтологии объектов.

Описанное выше взаимодействие онтологии ПдО и онтологии ПрП является одной из основных характеристик онтолого-управляемости ЗОИС, другие же характеристики приведены в [Палагин, 2009, б].

Онтолого-управляемая архитектура ИС

В многочисленных зарубежных публикациях под *онтолого-управляемой архитектурой* (Ontology-driven architecture) понимают такую архитектуру ЗОИС, в которой база знаний ПдО представлена формальной онтологией. При этом последняя описана на некотором языке описания онтологий, а список таких языков определён консорциумом W3C (OWL, RDFS и др.). ЗОИС с онтолого-управляемой архитектурой классифицируют [Палагин, 2005] по функциональному признаку на онтолого-знающие информационные системы (ОЗнИС) и онтолого-управляемые информационные системы (ОУИС). ОЗнИС характеризуется “пассивной” ролью онтологии, т.е. выполнением ею роли только информационно-справочного компонента. Любая другая подсистема только “знает” о существовании (возможно отдалённо) онтологии и может использовать её для любой специфической цели приложения. В ОУИС онтология играет “активную” роль, кооперируется с другими подсистемами в текущем времени в соответствии с “высшей” общей целью ЗОИС. Важным следствием использования онтологии в текущем времени является способность коммуникации между программными агентами. Последние поддерживают связь друг с другом через сообщения, которые содержат выражения, сформулированные в терминах онтологии (*онтолого-управляемая коммуникация*). Для того чтобы программному агенту “понять” значения этих выражений, ему необходим доступ к онтологии.

Известна и другая трактовка онтолого-управляемой архитектуры. Это новая концепция разработки программ – MDD (Model Driven Development – разработка, управляемая моделями) и вариант ее реализации – MDA (Model Driven Architecture – архитектура, управляемая моделью, или модельно-управляемая архитектура), предполагающая разработку архитектуры программного комплекса, ориентированную на предметную область и не зависящую от конкретной платформы программирования. Развитием данной концепции является подход к разработке программного обеспечения, предлагаемый консорциумом W3C для разработки, в частности, серверных приложений и называемый архитектурой, управляемой онтологией. [www.w3.org].

Выше приведено общепринятое (по крайней мере, в зарубежных публикациях) описание онтолого-управляемой архитектуры для ИС различного назначения. Для ЗОИС с обработкой ЕЯО оно является фундаментальным, основано на свойстве активности знаний, в том числе онтологических. Специфика обработки языковых и предметных знаний предполагает включение в архитектуру ОУИС соответствующих онтологий, которые в свою очередь упорядочиваются в соответствии с онтологической иерархией. И главной особенностью архитектуры ОУИС, подчёркивающей её онтолого-управляемость, является явное представление процедурных знаний в виде онтологии задач [Палагин, 2009, а].

На рис. 2 представлена схема фрагмента архитектуры ОУИС, отображающая только её “онтологическую” компоненту. Это “ядро” системы, между которым устанавливаются эксплицитные связи с процессами и процедурами (или технологической цепочкой) ИИТ обработки знаний, содержащихся в ЕЯО, с приложениями (через онтологию задач), с научной (онтология верхнего уровня) и языковой (языково-онтологическая) картинами мира. На нём приняты следующие обозначения: $ОД_k$ ПдО – онтология k -го домена предметных областей, где $k = \overline{1, K}$, $K = Card OD$ – мощность множества онтологий доменов, интегрированных в библиотеку; $О ПдО_i$ – онтология i -ой предметной области, где $i = \overline{1, I}$, $I = Card ПдО$ – мощность множества онтологий предметных областей, интегрированных в библиотеку; $ОЗ_j^i$ – онтология j -го типового набора задач i -ой предметной области, где $j = \overline{1, J}$, $J = Card OZ$ – мощность множества типовых наборов задач i -ой предметной области;

ЛО ПдО_{*i*} – лингвистическая онтология *i*-ой предметной области; ЯОКМ – языково-онтологическая картина мира; 1 – обработка омонимичных форм; 2 – построение поверхностных семантических структур; 3 – обработка семантических неоднозначностей; 4 – обработка концептуальных структур; ИИТ – интегрированная информационная технология обработки языковых и предметных знаний.

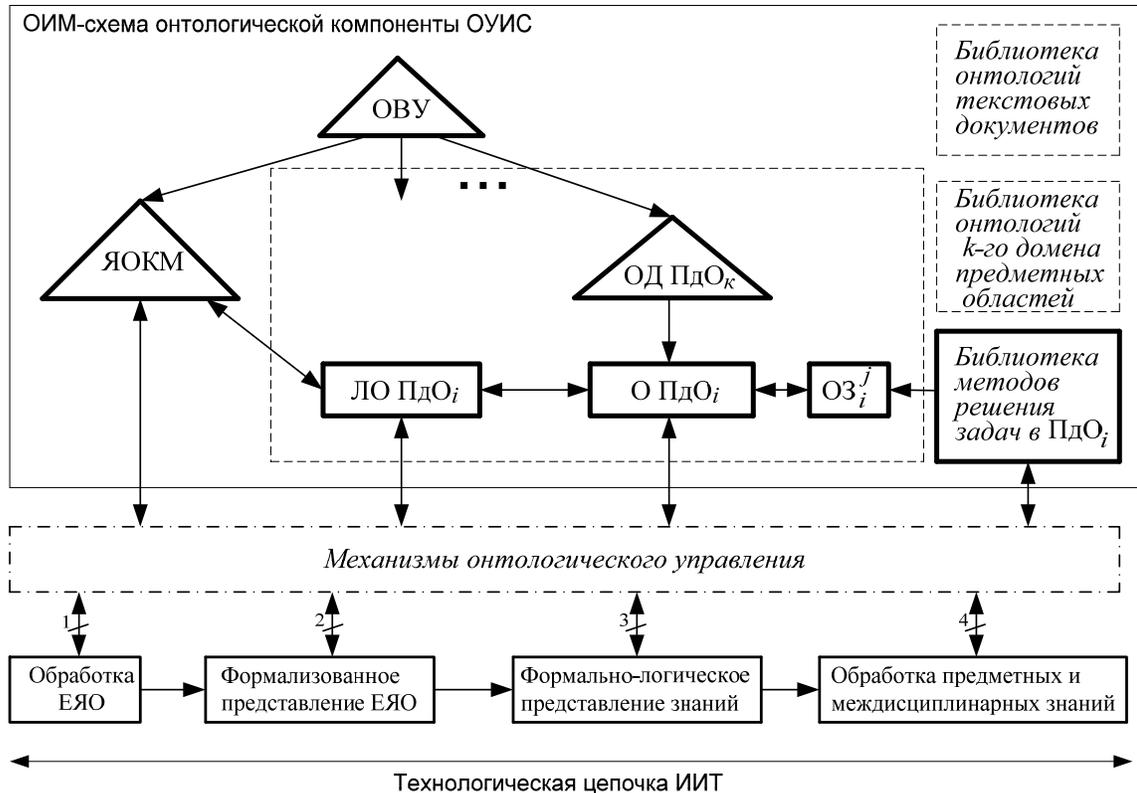


Рис. 2. Фрагмент архитектуры “онтологической” части ОУИС с обработкой ЕЮ

Итак, ЗОИС с онтолого-управляемой архитектурой присущи следующие характерные черты:

композиция онтологий разного уровня и назначения, как по вертикали, так и по горизонтали. По вертикали интегрируются онтология верхнего уровня, онтология домена предметных областей и онтология предметной области. По горизонтали интегрируются онтологии предметных знаний и знаний проблемного пространства. В свою очередь, предметные знания взаимодействуют с языковыми знаниями через лингвистическую онтологию предметной области;

- эффективное многократное использование онтологии предметной области и онтологии задач для разных наборов типовых задач. Будучи один раз построенными указанные онтологии пригодны для решения произвольного набора задач. При этом для новой задачи только составляется её расширенная спецификация на естественном языке, которая затем анализируется лингвистическими средствами ОУИС с обработкой ЕЮ. Результат анализа передаётся в блок решателя задач, где формируются обобщённый алгоритм решения задачи, наборы понятий-объектов и понятий-процессов, участвующих в решении задачи, и выбираются методы решения задачи;
- следствием композиции принципов и механизмов двойной парадигмы предметных знаний и онтологического управления является интеграция и эффективное применение компонентов архитектуры ОУИС в архитектурно-структурную организацию **интегрированных инструментальных средств автоматизированного построения онтологических баз знаний предметных областей**;

- частным случаем описанного в предыдущем пункте следствия является следующее. Произвольная онтология (простая, смешанная или формальная) эффективна, прежде всего, в обучении. Действительно, гораздо эффективнее показать новому сотруднику предприятия онтограф структурных подразделений предприятия, чем долго рассказывать о той же самой структуре. И только ИС с онтолого-управляемой архитектурой, имеющая в своём составе формальную онтологию некоторой предметной дисциплины, позволяет реализовать все преимущества электронного образования, в том числе и дистанционного обучения. Наиболее полно указанные свойства проявляются при использовании и взаимодействии онтологий двух уровней – онтологии домена предметных дисциплин (это может быть множество дисциплин, читаемых на кафедре или факультете) и онтологий самих предметных дисциплин. При этом естественным образом находят своё решение такие проблемы как: открытости и закрытости систем обучения; единообразия представления понятийных структур обучения; автоматизированного построения новых курсов обучения с учётом накопленных преподавателями кафедры знаний; замена кадрового состава преподавателей и др.

Выводы

В работе рассмотрены основные принципы парадигмы компьютерных онтологий, описана схема онтологий-компонент предметной области и онтологическая часть архитектуры знание-ориентированной информационной системы с обработкой ЕЯО, обладающая признаками онтолого-управляемости. Дальнейшее направление исследований связано с решением одной из главных научно-технических задач разработки ЗОИС с обработкой ЕЯО – построение инструментальных средств автоматизированного проектирования онтологии ПДО на основе больших объемов текстовой информации.

Литература

- [Аверкин, 1992]. Толковый словарь по искусственному интеллекту / Авт.-сост. А.Н. Аверкин, М.Г. Гаазе-Рапопорт, Д.А. Поспелов. – М.: Радио и связь, 1992. – 256 с.
- [Гаврилова, 2001]. Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. – СПб.: Питер, 2001. – 384 с.
- [Методический комплекс]. Методический комплекс по дисциплине “Современные проблемы науки” // Методология анализа предметных знаний. – Доступно на http://old.ulstu.ru/people/SOSNIN/umk/Modern_Scientific_Problems/beloborodov/item_znan.htm.
- [Палагин, 2005]. Палагин А.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: УНІВЕРСУМ. – 2005. – 680 с.
- [Палагин, 2009, а]. Палагин А.В., Петренко Н.Г. Системно-онтологический анализ предметной области // УСиМ. – 2009. – № 4. – С. 3–14.
- [Палагин, 2009, б]. Палагин А.В., Крывий С.Л., Петренко Н.Г. Знание-ориентированные информационные системы с обработкой естественно-языковых объектов: основы методологии и архитектурно-структурная организация // УСиМ. – 2009. – № 3. – С. 42–55.
- [Широков, 2005]. Широков В.А., Бугайов О.В., Грязнухина Т.О. та ін. Корпусна лінгвістика. – К.: Довіра, 2005. – 471 с.
- [Guarino, 1998]. Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.) Formal Ontology and Information Systems. //Proceedings of FOIS'98. – Trento, Italy. – 1998. – 6–8 June. – IOS Press, Amsterdam. – PP. 3–15.
- [Gruber, 1993]. Gruber, T. R. 1993. A translation approach to portable ontology specifications // Knowledge Acquisition, 5 (2): PP. 199–220.
- [www.w3.org]. <http://www.w3.org/2001/sw/BestPractices/SE/ODA/>.

Информация об авторе



Николай Петренко – старший научный сотрудник, Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40; e-mail: petrng@ukr.net

Область исследований: методология и инструментальные средства автоматизированного проектирования онтологий предметных областей, системная интеграция трансдисциплинарных научных знаний