

К ВОПРОСУ ПРОЕКТИРОВАНИЯ ОНТОЛОГОУПРАВЛЯЕМОЙ ИС ОБРАБОТКИ ЕЯО

Александр Палагин, Николай Петренко

Аннотация: В статье рассмотрен формальный подход и основное содержание методологии формализованного проектирования

Ключевые слова: онтолого-управляемой информационной системы обработки знаний, содержащихся в естественно-языковых объектах (ЕЯО).

ACM Classification Keywords: I.2.7 Natural Language Processing - Language parsing and understanding

Conference: The paper is selected from XIVth International Conference "Knowledge-Dialogue-Solution" KDS 2008, Varna, Bulgaria, June-July 2008

Постоянное увеличение объёмов научной информации в сети Интернет и других источников с естественным способом представления требует усовершенствования известных и разработки новых, более эффективных формальных подходов и моделей обработки, начиная от поиска релевантной информации, её анализа, устранения разного рода неоднозначностей, формально-логического представления, онтолого-семантического представления, информационно-кодowego представления, представления на формальном языке описания знаний, адаптации известных и разработки новых процедур работы со знаниями и соответствующих им алгоритмов и до получения конкретных результатов пользователя.

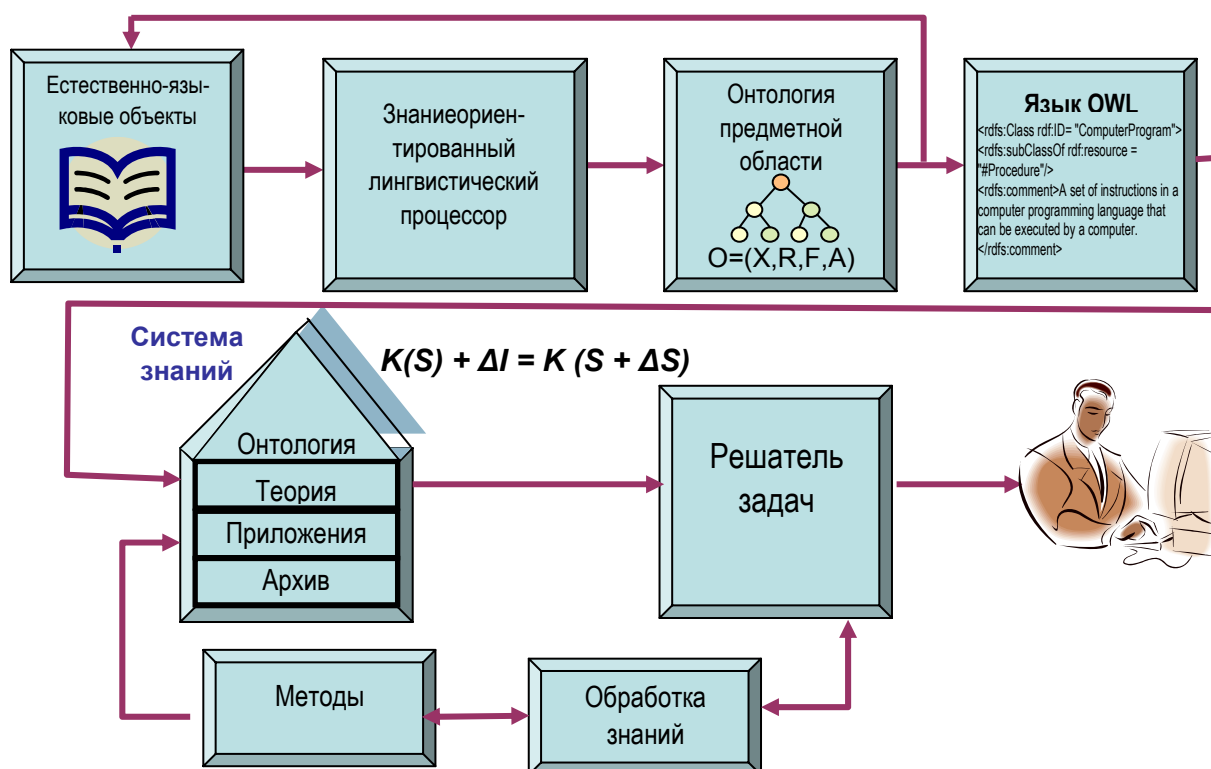


Рис.1. Компьютерная обработка знаний, содержащихся в естественно-языковых объектах

В более общем виде приведенная выше последовательность этапов обработки знаний, содержащихся в естественно-языковых объектах (ЕЯО) представляется цепочкой технологий Natural Language Processing (NLP) → Knowledge Representation (KR) → Knowledge Processing (KP) и показана на рис.1.

Проектирование компьютерных средств, реализующих объединение указанных технологий с учётом специфики входной информации (различного вида источники ЕЯО) можно представить как проектирование сложной системы, включающей языково-онтологическую информационную систему (ЯОИС), полную онтологию предметной области (ПрО) и систему обработки знаний в заданной ПрО. Архитектурно-структурная организация указанных подсистем, реализующая схему обработки знаний (рис.1), соответствует онтологууправляемой ИС обработки знаний, содержащихся в ЕЯО [1, 2].

Особенности архитектуры и структуры современных "онтологизированных" знаниеориентированных систем (или онтологууправляемых ИС) полностью определяют подход к их проектированию. Эти особенности можно разделить на системные и технологические.

Системные особенности вытекают из требования построения на базе онтологууправляемой ИС инструментального комплекса обработки знаний в заданной предметной области с различными техническими характеристиками, в том числе с возможностью подключения аппаратных средств поддержки, спроектированных на базе современных ПЛИС-технологий, потребительскими свойствами и сводятся к созданию класса проблемно-ориентированных онтологууправляемых ИС.

Технологические особенности определяются состоянием развития элементарно-технологической базы современных компьютеров и связаны, прежде всего, с появлением сверхмощных программируемых логических интегральных схем (ПЛИС), вызвавших интенсивное развитие новых методов проектирования знаниеориентированных ИС. Это объясняется целым рядом преимуществ совместного использования программных и аппаратных средств, составляющих в целом информационную систему [3].

В качестве фундаментального механизма при разработке средств интерпретации ЕЯО предлагается модификация известного логико-информационного подхода [4], применительно к цепочке технологий NLP → KR → KP, названном *онтолого-инфологическим* (ОИП), сущность которого состоит в следующем.

В соответствии с логической концепцией онтолого-инфологической модели компонентные процессоры представляются известной композицией сетей операционных и управляющих автоматов. Последняя моделирует иерархическую систему управления современных ИС обработки знаний, и, в свою очередь, описывается композицией программируемых автоматов.

В соответствии с информационной концепцией компонентные процессоры рассматриваются как информационная система, вся информация в которой отнесена к трём "сферам" состояний: хранения, транспортировки и преобразования. Очевидно, что при определённых соотношениях между объектами информации в этих сферах можно получить оптимальные технические параметры онтологууправляемой ИС.

Под уровнем управления $\tau_i (\forall i = 0 \div h)$ понимается совокупность структурных и информационных (программных) средств представления и интерпретации операторов языка соответствующего уровня системы, а также формирования необходимой последовательности их выполнения в соответствии с заданным алгоритмом. Наиболее известными и распространёнными в архитектурах современных процессоров и ЭВМ являются микропрограммный, программный и алгоритмический уровни.

Таким образом, формальное описание модели можно представить в виде $\overline{\bigvee_{i=0,h} \tau_i} = \langle A_i, \Lambda_i, R \rangle$, где A_i -

множество алгоритмов, реализуемых и записанных в памяти на i -м уровне; Λ_i - множество операторов программирования i -го уровня, в терминах которых представлены эти алгоритмы; R - множество информационно-кодовых представлений операторов i -го уровня.

Оптимальной считается такая структурная реализация модели онтологоуправляемой ИС, для которой в соответствии с принятыми критериями найдены оптимальное количество уровней и оптимальные соотношения между определёнными характеристиками компонент на каждом уровне, а также соответствующими характеристиками компонент соседних уровней.

Кратко рассмотрим методологию проектирования указанной онтологоуправляемой ИС, которая в свою очередь представляется композицией методик формализованного проектирования подсистем обработки ЕЯО, полной онтологии ПрО и обработки знаний.

Подсистема обработки ЕЯО представляет собой ЯОИС, одной из отличительных особенностей которой от известных лингвистических процессоров является наличие языково-онтологической картины мира (ЯОКМ), которая, по сути, представляет собой онтологию для предметной области “компьютерная обработка ЕЯ”. Отметим, что подход к проектированию ЯОИС и ЯОКМ, их структурная и информационная модели рассмотрены в ряде работ [1, 5-9].

Подсистема полной онтологии ПрО представляет собой системную интеграцию онтологических знаний проблемной области и естественно-языкового обеспечения [10], что в общем случае представляется для первых – как композиция онтологии ПрО и одной из известных онтологий верхнего и среднего уровней (например, SUMO, онтология Дж. Совы или Mikrokosmos), а для вторых – композицией ЯОКМ и лингвистической онтологии для заданной ПрО. Приведём этапы и подэтапы проектирования полной онтологии ПрО, из названий которых следует смысл выполняемых работ на соответствующем этапе.

Этап составления технического задания (ТЗ) на проектирование полной онтологии ПрО включает: изучение и систематизацию начальных условий; определение перечня и краткое описание “активных” и “пассивных” процедур (эти понятия соответствуют введенным в [2] определениям “онтологоуправляемой” и “онтологознающей” ИС), что выполняются в блоке онтологии ПрО, в том числе процедуры системной интеграции в библиотеку онтологий; определение формальной теории представления онтологии ПрО на некотором языке L ; определение схемы апробации разработанной онтологии ПрО.

Следующим, *основным этапом проектирования онтологии ПрО* является построение множеств $\{X\}$, $\{R\}$, $\{F\}$, $\{A\}$, представляющих кортеж известной схемы формальной модели онтологии ПрО. При этом систематизируются результаты предыдущего этапа, включающего, в том числе анализ, группирование и фиксацию знаний о предметной области T .

Обобщённая последовательность шагов алгоритма реализации основного этапа проектирования онтологии ПрО включает следующее.

Определение множеств $\{X\} = \{x_i, i = \overline{1, N}\}$, где N – множество концептов из T , и $\{A\} = \{a_j(d_l, r_{S_m})\}$, $j = \overline{1, J}$, $l = \overline{1, L'}$, $m = \overline{1, M}$, где J, L', M – соответственно мощность множеств аксиом, определений и ограничений для X .

Определение множества $\{R\} = \{r_p\}$, $p = \overline{1, P}$, где P – мощность множества концептуальных отношений в T и их ранжирование по степени важности (т.е. построение кортежа $\langle r_1, r_2, \dots, r_p \rangle$, где индекс 1 присвоен наиболее существенному концептуальному отношению в T , а p – наименее существенному).

Заметим, что формирование базовых наборов множеств $\{X\}$, $\{A\}$ и $\{R\}$ осуществляется оригинальными инструментальными средствами, названными соответственно “Concepts and Axioms Formation” и “Relations Formation”.

Разработка функционально-ориентированных компонентов онтологии ПрО.

3.1. Адаптация онтологии верхнего и среднего уровня к онтологии заданной ПрО.

3.2. Разработка лингвистической онтологии предметной области.

Определение множества $\{F\} = \{f_q\}, q = \overline{1, Q}$, где Q – мощность множества функций интерпретации, на кортеже $\langle X, R \rangle$, оптимальных для T .

Графическое проектирование компонентов онтологии у выбранной автоматизированной инструментальной среде и их системная интеграция в полную онтологию ПрО.

Создание файла формального описания разработанной онтологии ПрО.

Для данного этапа существенно не только разработать максимально полные множества, входящие в кортеж схемы формальной модели онтологии, но и обеспечить автоматическую проверку на целостность и непротиворечивость совокупности элементов упомянутых выше множеств.

Этап *апробации разработанной онтологии* полностью зависит от содержания подэтапа ТЗ, на котором разрабатывалась схема апробации для формальной модели. Теперь эту схему необходимо реализовать на реальном прототипе онтологии. В дополнение к указанной схеме необходимо выбрать из построенного множества функций интерпретации $\{F\}$ такие функции, которые позволили бы выполнить всестороннее тестирование. При этом особое значение имеют тестовые последовательности. По результатам тестирования разработчику необходимо рассмотреть вопрос о необходимости повторения некоторых шагов проектирования.

Подсистема обработки знаний (ПОЗ) в общем виде представляется структурой, представленной на рис.2.

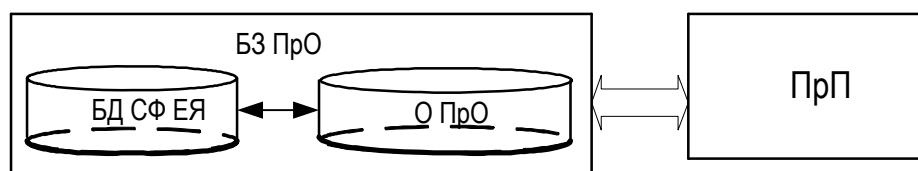


Рис.2. Обобщённое структурное представление ПОЗ

На рисунке приняты следующие обозначения:

БЗ ПрО – база знаний предметной области;

БД СФ ЕЯ – база данных суждений и фактов из заданной ПрО на естественном языке;

О ПрО – онтология ПрО;

ПрП – прикладной процессинг.

Особенности заданной ПрО полностью определяют подход к проектированию компонент подсистемы обработки знаний – БЗ ПрО и алгоритмов ПрП.

Для компоненты БЗ ПрО необходимо создать базу данных суждений и фактов и разработать алгоритмы её взаимодействия с онтологией ПрО, спроектированной в подсистеме полной онтологии ПрО. Указанное проектирование выполняется в соответствии с известными методами.

Проектирование компоненты ПрП в общем случае представляется разработкой совокупности алгоритмов, реализующих заданное множество задач пользователя, и их взаимодействия с БЗ ПрО и результатами распознанных знаний (полученными в подсистеме ЯОИС), содержащихся во входном ЕЯО. При этом аппаратная реализация указанных алгоритмов осуществляется в соответствии с методологией САПР ПЛИС, описанной в [3].

В заключение отметим, что полная онтология ПрО в процессе реализации цепочки технологий играет различную роль, а точнее – активны её различные компоненты.

При NLP-технологии основную роль играет внутриязыковое обеспечение ПрО – интеграция лингвистической онтологии из соответствующим фрагментом ЯОКМ. Оно поддерживает выполняемые ЯОИС процедуры анализа и распознавания знаний, содержащихся во входном ЕЯО.

При KR–технологии выполняется переход к формальному описанию знаний на некотором языке L , при котором активную роль играют онтологические знания ПрО, представленные онтографом. При этом формируются онтограф и интерпретационная структура, соответствующие знаниям, содержащимся во входном ЕЯО.

При КР–технологии активную роль играет база знаний ПрО, в состав которой входит формальное описание онтологических знаний и её аксиоматизация.

Литература

1. Палагин А.В., Яковлев Ю.С. Системная интеграция средств компьютерной техники. – Винница: «УНІВЕРСУМ-Вінниця», 2005. – 680 с.
2. 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.
3. Реконфигурируемые вычислительные системы: Основы и приложения. / А.В. Палагин, В.Н. Опанасенко. – К.: Просвита, - 2006. – 280с.
4. Палагин А.В. К решению основной задачи эмуляции // УСиМ. – 1980. – №3. – С.24-28.
5. Палагин А.В. Организация и функции "языковой" картины мира в смысловой интерпретации ЕЯ - сообщений//Information Theories and Application. – 2000. – Vol. 7, №4. С.155-163.
6. Палагін О.В., Петренко М.Г. Модель категоріального рівня мовно-онтологічної картини світу //Математичні машини і системи. – 2006. - №3. - С.91-104.
7. Палагін О.В., Петренко М.Г. Архітектурно-онтологічні принципи розбудови інтелектуальних інформаційних систем //Математичні машини і системи. – 2006. - №4. - С.15-20.
8. Палагін О.В., Петренко М.Г. Розбудова абстрактної моделі мовно-онтологічної інформаційної системи //Математичні машини і системи. – 2007. - №1. - С.42-50.
9. Палагин А.В. Архитектура онтологоуправляемых компьютерных систем //Кибернетика и системный анализ. 2006 - №2. – С.111-124.
10. Палагин А.В., Петренко Н.Г. К вопросу системно-онтологической интеграции знаний предметной области //Математические машины и системы. – 2007. – №3,4. – С.63-75.

Информация об авторах

Палагин Александр Васильевич - Ин-т кибернетики им. В.М. Глушкова НАН Украины, Киев-187 ГСП, 03680, просп. акад. Глушкова, 40, e-mail: petrng@ukr.net

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