

---

---

## RULE-MINING: ПОДХОД К АВТОМАТИЗИРОВАННОМУ ИЗВЛЕЧЕНИЮ ОНТОЛОГИЙ

Роман Гадиатулин, Светлана Чуприна

**Аннотация:** В статье описывается подход к реализации автоматизированного извлечения онтологий из продукционных баз знаний, вопросы организации извлечения, хранения и использования онтологий для широкого круга проблемных областей. Разработанная в рамках предлагаемого подхода оболочка *OntoMaker 2.0*, наряду с возможностями автоматизированного извлечения знаний предоставляет визуальные средства для редактирования и построения онтологий, а также широкие возможности расширения и настройки.

**Keywords:** извлечение знаний, автоматическая каталогизация и классификация, онтологический подход, проектирование и разработка интеллектуальных информационных систем.

**ACM Classification Keywords:** 1.2 Artificial Intelligence: 1.2.7 Natural Language Processing – Text analysis; D.2 Software Engineering: D.2.2 Design Tools and Techniques – Computer-aided software engineering (CASE).

---

### Введение

---

Современное состояние дел в отдельных областях компьютерных наук диктует необходимость привлечения методов инженерии знаний для решения широкого класса практических задач. Ярким примером тому является инициатива Semantic Web, основная цель которой – наделить огромные массивы данных, опубликованных в сети Internet большей осмысленностью, повысить удобство работы с этой информацией. Одним из главных достижений проекта Semantic Web стала разработка стандарта описания онтологий – OWL (Ontology Web Language), благодаря чему множество инженеров по знаниям, программистов и экспертов получили возможность использовать общие правила представления, хранения и обработки онтологий.

Существуют различные толкования самого понятия онтологии [Гаврилова,2000]. В данной работе под онтологией понимается структурная спецификация некоторой предметной области, ее концептуальное описание в виде формализованного представления, которое включает словарь терминов предметной области и логические выражения, описывающие взаимосвязи этих понятий. Таким образом, онтология некоторой предметной области представляет собой тезаурус понятий этой предметной области, обеспечивающий возможность толкования терминов предметной области посредством интерпретации таких типов парадигматических отношений как «часть-целое», «класс-подкласс» и некоторых видов ассоциативных связей.

На волне интереса к онтологиям были созданы инструментальные средства и механизмы, специально ориентированные на широкое применение онтологий в задачах интеллектуального поиска, классификации, выявления несогласованности в данных, моделирования поведения интеллектуальных агентов. Однако даже наличие хорошего инструментального окружения не снимает проблем, связанных с трудностью проектирования и построения самих онтологий, а автоматизация процесса извлечения онтологий, как и в целом, задача извлечения знаний, и по настоящее время не имеют своего

эффективного решения. Тем ценнее становятся уже разработанные онтологии и опыт их использования для решения широкого круга задач.

Кроме того, к настоящему времени накоплен большой объем баз знаний (БЗ), созданных не только в рамках различных инструментальных сред, но и на основе различных парадигм представления знаний (продукционной, логической, фреймовой, на семантических сетях). В процессе создания современных интеллектуальных информационных систем зачастую требуется интеграция знаний из разнородных источников и, как следствие, эффективное решение задач, связанных с тиражированием знаний. По-прежнему не имеет своего удовлетворительного решения проблема автоматизации процесса выбора адекватного специфике конкретной проблемной области и принятого в ней стиля рассуждения экспертов средства представления знаний. Поэтому и по сей день актуальны исследования, направленные на разработку такого подхода к представлению и тиражированию знаний, который с одной стороны позволял бы наиболее адекватно учитывать специфику проблемной области, а с другой – представлять и использовать знания в некотором унифицированном виде.

В данной статье мы не касаемся вопросов, связанных с решением указанных проблем на основе применения современных мультиагентных технологий. Нам представляется весьма актуальным применение онтологий для решения перечисленных выше проблем в рамках интеллектуальных информационных систем, основанных на использовании баз знаний традиционных экспертных систем (ЭС), которые по-прежнему остаются одними из самых распространенных приложений искусственного интеллекта на практике. Предлагаемый нами в рамках проекта XG# подход к извлечению онтологий проблемных областей предусматривает автоматизацию процесса анализа уже существующих продукционных баз знаний (rule-mining) и выявления в них скрытых зависимостей, которые могут быть явно представлены в виде основных парадигматических отношений. Построенная таким образом онтология в значительной мере способствует обнаружению несогласованностей и противоречий в исходной БЗ, а также облегчает ее сопровождение и отладку. Кроме того, благодаря поддержке стандарта OWL, обеспечивается возможность тиражирования знаний между различными когнитивными системами, что автоматизирует процесс создания и пополнения баз знаний этих систем.

---

## Проект XG#

---

XG# – это инструментальная среда разработчика оболочек экспертных систем с интегрированным представлением знаний, основанная на декларативном описании грамматики языка представления знаний, учитывающего специфику конкретной проблемной области [Чуприна,2003-2006], [Гадиятулин,2006]. Описание грамматики конкретного языка представления знаний в среде специально разработанного визуального инспектора грамматик транслируется во внутреннее представление системы, на основе которого генерируются все основные компоненты оболочки ЭС (компоненты приобретения знаний, логического вывода, объяснения).

В основе предлагаемого подхода – активное использование метазнаний не только для описания синтаксиса и семантики языка представления знаний, но и для формирования онтологий, описывающих основные виды взаимосвязей между понятиями проблемной области, а также модель пользователя системы. Однако, как уже отмечалось во введении, проектирование и разработка онтологий, то есть онтологический инжиниринг, не является тривиальной задачей. Он требует от разработчиков профессионального владения технологиями инженерии знаний — от методов извлечения знаний до их структурирования и формализации [Гаврилова,2000].

Перечислим несколько важных, на наш взгляд, задач, решение которых можно осуществить при помощи привлечения онтологий за счет явного представления и учета основных парадигматических отношений между понятиями проблемной области:

- упрощение отладки и модификации базы знаний ЭС за счет визуализации связей между понятиями, представленными в БЗ системы;
- обеспечение дополнительных знаний для автоматической генерации объяснений хода логического вывода;
- тиражирование знаний посредством публикации онтологий, полученных с помощью автоматизации процесса извлечения скрытых знаний из баз знаний уже существующих ЭС или созданных экспертом «с нуля» в среде визуального редактора онтологий;
- ускорение процесса разработки ЭС за счет автоматизации этапа первоначального заполнения БЗ на основе существующих в данной проблемной области онтологий;
- трансформация и приведение к единому виду знаний из БЗ различных ЭС.

### Методика rule-mining

В проблематике искусственного интеллекта большой популярностью в настоящее время пользуются подходы, основанные на автоматизации извлечения знаний из данных, текстов и веб-ресурсов (data-mining, text-mining, web-mining). В рамках проекта XG# мы исследовали возможность автоматизации извлечения знаний более высокого уровня из предметных баз знаний производственных ЭС. Такой подход можно условно назвать rule-mining. Исходными данными для rule-mining служит текст базы знаний, из которого автоматически извлекаются закономерности, которые в явном виде представляются в среде визуального редактора онтологии для целей дальнейшего анализа и редактирования пользователем, что в значительной мере упрощает процесс построения онтологии проблемной области. В результате реинжиниринга из БЗ автоматически извлекаются дерево целей и описание доменов значений понятий, представленных в ЭС. Процесс извлечения онтологии схематично представлен на рис.1.

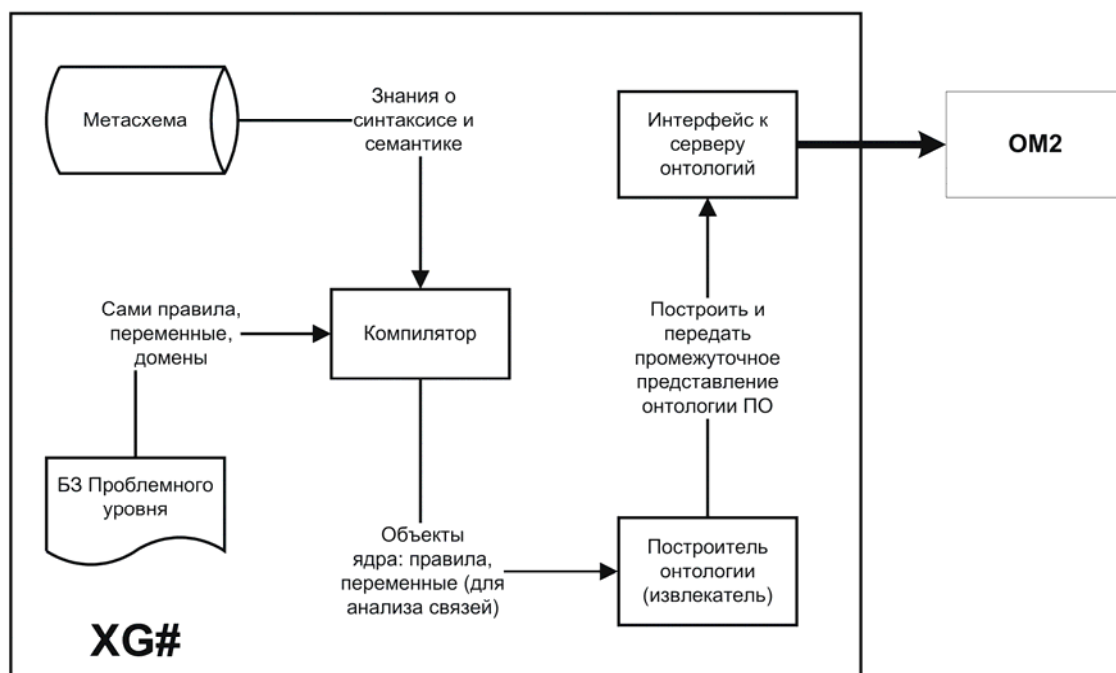


Рис. 1. Процесс извлечения онтологии

Построитель онтологии анализирует возможность наличия взаимосвязей между понятиями БЗ на основе некоторых эвристик. Например, в результате анализа фактов, содержащихся в посылке и заключении правила, система может сделать заключение о наличии между ними связи типа «подкласс-класс». Далее автоматически на основе метазнаний генерируются наборы вопросов пользователю-эксперту с целью уточнения семантики выявленных взаимосвязей, анализируются ответы пользователя и строится соответствующая онтология. Результирующая онтология таким образом в явном виде отражает различные виды взаимосвязей (например, «класс-подкласс», «часть-целое») понятий проблемной области, которые неявно содержатся в фактах и правилах исходной БЗ предметного уровня. Далее онтология автоматически сохраняется в виде описания классов, экземпляров и свойств в нотации OWL и может быть с успехом использоваться для решения перечисленных выше задач.

Результирующая онтология, возможно, является неполной и неточной, однако эксперт имеет мощное средство для ее визуального редактирования и пополнения.

---

### **Визуальный инструментарий для работы с OWL онтологиями OntoMaker 2.0 (OM2)**

---

На сегодняшний день для большинства инструментальных средств построения онтологий характерно следующее. Во-первых, большая часть подобных систем хоть и имеет визуальную составляющую, однако некоторые конструкции приходится набирать вручную, что повышает уровень требований к разработчику онтологий – перед тем, как приступить непосредственно к своей работе, эксперт вынужден тратить время на изучение языка представления знаний. Во-вторых, часть инструментальных средств реализуют определенную функциональность для выполнения запросов к онтологиям, но, к сожалению, не имеют унифицированного интерфейса для формирования и выполнения запросов из внешних приложений. В-третьих, практически нет свободно распространяемых и ориентированных на конечного пользователя редакторов онтологий, что, естественно, замедляет развитие всего направления онтологического инжиниринга.

При разработке инструментальной среды разработчика онтологий OM2 мы постарались нивелировать минусы аналогов и перенять их достоинства. В OM2 любой объект онтологии имеет графическое представление (не только классы и индивиды, но и свойства, связи и др.). OM2 – независимое приложение, которое способно выступать в качестве сервера онтологий. В настоящее время редактор полностью поддерживает конструкции диалекта Lite языка описания онтологий OWL, ведется работа по расширению ее возможностей до диалекта DL. На рис. 2 приведена архитектура OM2.

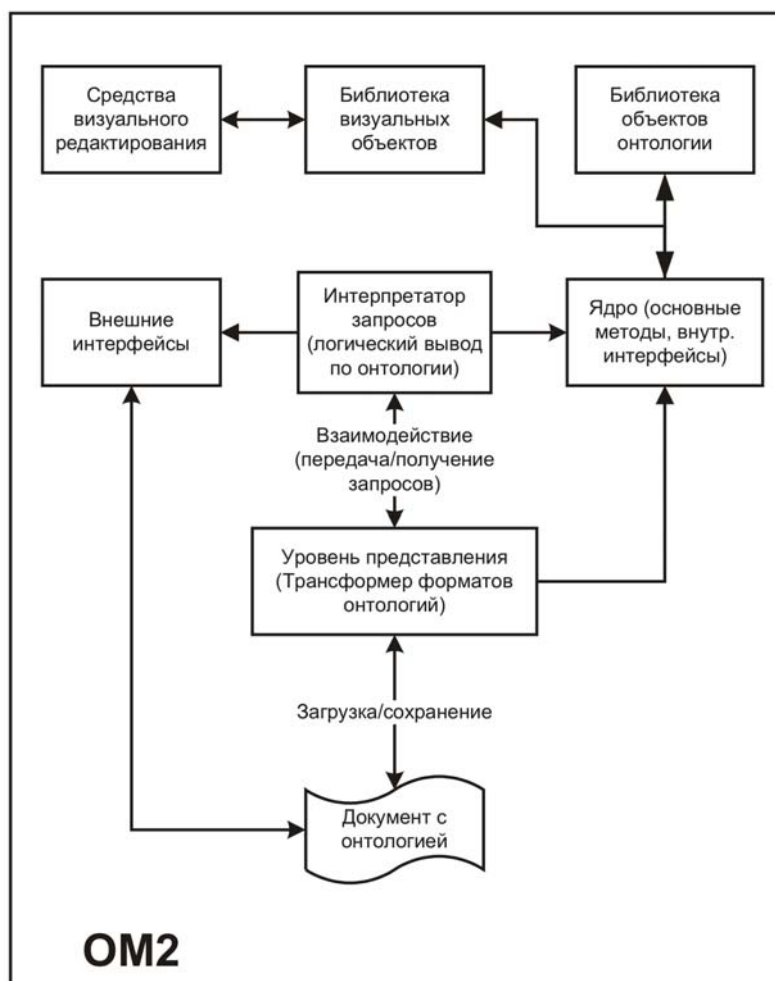


Рис. 2. Архитектура OM2

В заключение отметим, что описанный подход прошел апробацию на нескольких международных конференциях (см. [Чуприна,2004-2006], [Гадиагулин,2006]). Обнадеживающие результаты были получены при использовании OntoMaker версии 1.0 для автоматизированного извлечения онтологий из БЗ экспертной системы по определению химических веществ, разработанной в Пермском государственном университете, и ряда других ЭС, находящихся на стадии исследовательского прототипа, что подтверждает жизнеспособность предлагаемого подхода.

На рис.3 приведен фрагмент онтологии, автоматически построенной в разрабатываемой нами инструментальной среде на основе анализа базы знаний демонстрационной ЭС по определению химических веществ.

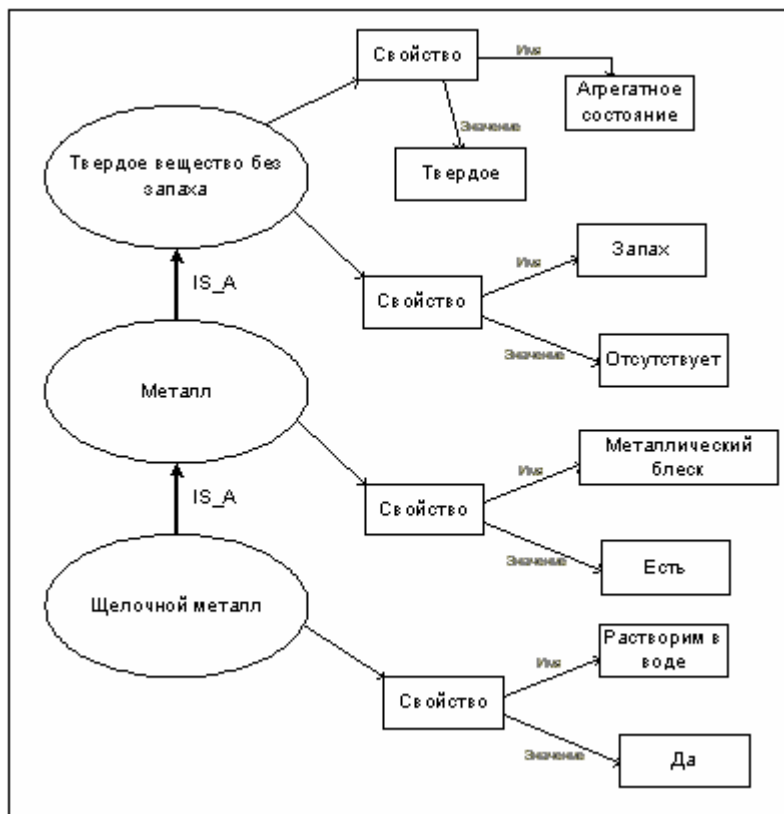


Рис. 3. Фрагмент онтологии базы знаний демонстрационной ЭС по определению химических веществ

## Результаты

Область применения онтологий, построенных вручную с помощью визуального инструментария OM2, внешних онтологий, опубликованных в Web, или онтологий, извлеченных посредством rule-mining, очень широка: от обнаружения и исправления ошибок при вводе (модификации) знаний до автоматической генерации на естественном языке осмысленных объяснений хода логического вывода ЭС. Сам процесс автоматизированного построения онтологий способствует более глубокому анализу проблемной области и упрощению отладки ЭС. Создав онтологию, эксперт облегчает работу себе, пользователю и, возможно, другим разработчикам экспертных систем в данной проблемной области. Само по себе существование онтологии, естественно, не решает описанных в данной статье проблем. Для этого онтология должна быть интегрирована в систему таким образом, чтобы было обеспечено ее активное и эффективное использование всеми компонентами экспертной системы, а также возможность ее дальнейшего наращивания и пополнения, в том числе посредством тиражирования знаний из внешних источников.

Построение онтологии не только цель, но и средство, облегчающее разработку ЭС:

Во-первых, при автоматическом формировании дерева целей упрощается отладка, появляется инструмент для контроля непротиворечивости знаний, наличия циклических ссылок, изолированных веток вывода и т.д. Эксперт и инженер по знаниям получают возможность проанализировать каузальные зависимости, сравнить выявленные взаимосвязи между понятиями БЗ экспертной системы с реально существующими в соответствующей проблемной области. Автоматически построенное дерево целей является хорошим дополнением к традиционным средствам отладки ЭС.

Во-вторых, на основе автоматизированного разбора БЗ, выявленные и представленные в явном виде знания о каузальных зависимостях облегчают реинжиниринг базы знаний. Занесение и удаление правил,

создание и удаление переменных становится легче контролировать: все изменения можно отслеживать по дереву целей и выдавать соответствующие предупреждения. Аналогичным образом упрощается задача миграции базы знаний в другую оболочку (например: из Guru в XG#).

В-третьих, возникает возможность разделения онтологий между экспертными системами для одной проблемной области, тиражирования наиболее удачных, общепризнанных онтологий, что влечет сокращение общего срока разработки ЭС или другой системы, основанной на тиражировании и использовании знаний.

---

### Библиографический список

---

- [Гаврилова,2000] Гаврилова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем: Учебник для вузов. — СПб: «Питер», 2000.
- [Чуприна,2003] Жигалов А.В., Никулин М.Б., Чуприна С.И. Оболочка экспертных систем XG#: концепция построения и реализация // Математика программных систем: межвуз. сб. научн. тр. / Перм. ун-т. Пермь, 2003. С. 97-106.
- [Чуприна,2004] Chuprina S., Lanin V., Borisova D., Khaeva S. Internet Intelligent Search System SmartFinder // Proc. of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology. Knowledge-Based Media Analysis for Self-Adaptive and Agile Multimedia Technology/ The Royal Statistical Society, November 25–26, 2004, London, U.K. P. 151–156.
- [Чуприна,2005] Chuprina S., Nikulin M. XG#: the Toolkit for Integrated Knowledge Representation Expert System Shells Development// Advances in Current Natural Sciences. Vol.5, 2005. Proc. of XXXII International Conference IT+S&E'2005. Ukraine, Crimea, Yalta-Gurzuff, 2005. P. 105.
- [Чуприна,Гадиатулин,2006] Чуприна С.И., Гадиатулин Р.А. XG# 1.0: Использование метазнаний для интеграции различных баз знаний //Актуальные проблемы математики, механики, информатики: материалы Международной научно-методической конференции, посвященной 90-летию высшего математического образования на Урале / Перм. гос. ун-т. Под ред. Лядовой Л.Н., Яковлева В.И., Ясницкого Л.Н. – Пермь, 2006. С. 104-105.
- [Чуприна,2006] Чуприна С. И. Интеллектуальная поисковая система SmartFinder: подход на основе онтологий // «Вестник Пермского университета», серия «Математика. Механика. Информатика», выпуск 4(4) / Перм. гос. ун-т. Пермь. 2006. С.118-122
- [Чуприна,Лядова,Ланин,2006] Чуприна С.И., Лядова Л.Н., Ланин В.В. Система интеллектуального поиска и автоматической каталогизации документов на основе онтологий // Proceedings of the XII-th International Conference “Knowledge-Dialogue-Solution” (KDS 2006), June 20-25, 2006, Varna (Bulgaria).- Sofia: FOI-COMMERCE, 2006, С. 139-144.
- [Гадиатулин,2006] Гадиатулин Р.А. Чуприна С.И. XG# 1.0 Реализация подхода к автоматическому построению онтологии из текстов баз знаний. Научно-технический, рецензируемый журнал общественного объединения Белорусской инженерной академии 1(21)/1 '2006 с. 39-42.

---

### Информация об авторах

---

**Роман Гадиатулин** – Пермский государственный университет, студент магистратуры кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: [gadiatuln@ramlber.ru](mailto:gadiatuln@ramlber.ru)

**Светлана Чуприна** – Пермский государственный университет, доцент кафедры математического обеспечения вычислительных систем; Россия, г. Пермь, 614990, ул. Букирева, д. 15; e-mail: [chuprinas@inbox.ru](mailto:chuprinas@inbox.ru)