

## Поддержка баз данных с онтологическими зависимостями на основе дескриптивных логик

Л. А. Калиниченко

leonidk@synth.ipi.ac.ru

Институт проблем информатики РАН  
Россия, г. Москва, 117333, ул. Вавилова, 44/2

**Аннотация.** В статье<sup>1</sup> дан краткий анализ состояния работ в области «онтологически базированных» систем доступа к данным и их возможного влияния на развитие информационных систем и баз данных<sup>2</sup>. Обсуждены вопросы соотношения онтологического и концептуального моделирования и соответствующих языковых средств. Дан анализ развития языков на дескриптивных логиках, ориентированных на использование в качестве средств концептуального моделирования в контексте баз данных и информационных систем. Дан краткий обзор известных экспериментальных результатов создания «онтологически базированных» систем доступа к данным.

**Ключевые слова:** онтологии, онтологические зависимости данных, концептуальное моделирование, дескриптивные логики, реляционные модели данных, язык Datalog, информационные системы.

### 1 Введение

В информатике под онтологией принято понимать формальное представление знаний в виде множества понятий некоторой предметной области и связей между такими понятиями. Такое представление используется для проведения рассуждений относительно сущностей предметной области, а также для ее описания. Таким образом, онтология представляет собой согласованный в сообществе словарь, который можно использовать для моделирования предметной области. Онтологии используются в различных областях информатики, включая: искусственный интеллект, Семантический Веб, системную инженерию, инженерию софтвера, биомедицинскую информатику, библиотечные науки, информационные архитектуры и др., как форма представления знаний о пространстве дискурса. Современные онтологии обладают большим структурным подобием, независимо от конкретного языка представления. Так, большинство онтологий определяют индивиды (экземпляры), классы (понятия), атрибуты и отношения. Примеры онтологий, разработанных для различных предметных областей (таких, как, например, культурное наследие, лингвистика, науки о Земле, геномика, систематика белков, фармацевтика, структура растений, образование, биология, архитектура предприятий, биомедицина, и др.) можно найти в [http://en.wikipedia.org/wiki/Ontology\\_\(information\\_science\)](http://en.wikipedia.org/wiki/Ontology_(information_science)). Эти примеры, а также развитие разнообразных языков онтологического моделирования свидетельствуют о практической востребованности онтологий для структуризации и представления знаний в различных предметных областях.

Последние годы отмечены активизацией экспансии онтологий в область информационных систем и баз данных. В основе такого распространения лежит простая идея – использовать

<sup>1</sup> Работа выполнена при частичной финансовой поддержке РФФИ (гранты 10-07-00342-а, 11-07-00402-а, и Программы фундаментальных исследований Президиума РАН №15, проект 4.2.

<sup>2</sup> Настоящая публикация является журнальным вариантом текста доклада Л.А.Калиниченко «Экспансия онтологий: «онтологически базированные» информационные системы», опубликованного в сборнике трудов Второго симпозиума «Онтологическое моделирование», Москва, ИПИ РАН, 2011 г.

онтологии как концептуальные схемы баз данных. В последнее время онтологические языки и системы довольно часто используются для представления и поддержки концептуальных схем над (реляционными) базами данных. Такая конструкция позволяет использовать аксиомы концептуальных схем и средства онтологических машин вывода при интерпретации запросов к базам данных. Этот подход практически используется при создании относительно небольших приложений, обычно в научных предметных областях. Подобная деятельность сопровождается введением ряда новых терминов – таких, как Онтологически Базированные Информационные Системы (Ontology Based Information Systems – OBIS), Онтологически Базированный Доступ к Данным (Ontology Based Data Access – OBDA), Онтологически Базированная Интеграция Данных (Ontology Based Data Integration - OBDI) [1,2,3,4]. Целью настоящей статьи является краткий анализ состояния работ в области «онтологически базированных» систем и анализ применяемых в них средств, основанных на дескриптивных логиках. Этот анализ выполняется на фоне работ по расширению реляционной модели данных специальными зависимостями, представленными в семействе языков Datalog $\pm$  [5,6,7,8,9,10,11,12]. Во главу угла такого расширения поставлено обеспечение поддержки онтологических зависимостей в базах данных на основе технологии баз данных. Однако, рассмотрение возможностей этих языков выходит за рамки данной работы, нацеленной на «онтологически базированный» доступ к данным на основе дескриптивных логик.

Статья организована следующим образом. В разделе 2 обсуждаются вопросы соотношения онтологического и концептуального моделирования и соответствующих языковых средств. В разделе 3 дан анализ развития языков на дескриптивных логиках, применяемых или специально разработанных для использования в качестве средств концептуального моделирования в контексте баз данных и информационных систем. В четвертом разделе приводится краткий обзор известных экспериментальных результатов создания «онтологически базированных» систем доступа к данным. Разделы 2 – 4 являются изложением рассмотренных в [13] вопросов, включающих анализ эволюции определений онтологии, места онтологических языков и, собственно, онтологических моделей в информатике. В заключении подводятся итоги выполненному в статье анализу.

## **2 Онтологии как концептуальные схемы**

Прежде всего, интересно проследить, как происходила эволюция определения понятия «онтология» в информатике. В начале 90-х годов прошлого века под воздействием усилий, направленных на создание стандартов интероперабельности компонентов информационных систем, онтологии были идентифицированы в качестве необходимого компонента систем, основанных на знаниях. Согласно определению Т.Грубера, данному в это время и широко используемому до сих пор, онтология есть спецификация концептуализации [14], или, иными словами, онтология представляет собой формальное определение понятий и их отношений в некоторой предметной области.

Еще в 1998 г. Н.Гуарино дал анализ возможных подходов создания информационных систем (ИС), движимых онтологиями, в котором предначертал варианты использования онтологий в процессе проектирования ИС и в процессе их функционирования [15]. Лейтмотивом этой работы является подчеркивание независимости спецификаций, заданных онтологиями, от каких-либо аспектов реализации. К сожалению, при этом не была прослежена параллель с ранее развитыми методами концептуального моделирования, семантически интероперабельных систем, систем интеграции информационных ресурсов, и др., неотъемлемым атрибутом которых является применение уровня спецификации, не зависящего от реализации.

Показательно также, насколько в последнее время развилось понятие «онтология» в сравнении с ранним определением Т.Грубера. Представительной является разделяемая многими более поздняя точка зрения все того же Т.Грубера [16], которую можно характеризовать следующими высказываниями. В контексте информатики, онтология определяет множество примитивов представления, при помощи которых осуществляется моделирование домена знаний или дискурса. Примитивами представления обычно являются классы (или множества), атрибуты (или свойства), и связи (или отношения) между экземплярами классов. В частности, в

контексте систем баз данных онтологии можно рассматривать как уровень абстракции моделей данных<sup>1</sup>, выраженных в иерархических и реляционных структурах, однако ориентированных на моделирование знаний об индивидах, их атрибутах и их связях с другими индивидами<sup>2</sup>. Говорят, что онтологии находятся на «семантическом» уровне, в то время как схемы баз данных являются моделями данных «логического» или «физического» уровня<sup>3</sup>. Ввиду их независимости от моделей данных ниже лежащих уровней, онтологии пытаются применять для интеграции неоднородных баз данных, обеспечения интероперабельности различных систем, а также для спецификации интерфейсов независимых сервисов, основанных на знаниях<sup>4</sup>.

Можно встретить также определения, подобные тому, что недавно было дано Д.Кальванезе [17]. Онтология является схемой представления, определяющей формальную концептуализацию предметной области. При этом спецификация онтологии обычно охватывает два различных уровня. Интенциональный уровень специфицирует множество концептуальных элементов и правил для определения концептуальных структур предметной области (в терминах дескриптивных логик говорят, что такие определения сосредоточены в TBox'e). Экстенциональный уровень задает множество экземпляров интенциональных элементов<sup>5</sup> (сосредоточенных в ABox'e).

Утверждается, что онтологии могут также использоваться как спецификации метауровня для описания модельных категорий, экземплярами которых являются концептуальные элементы. Наряду с анализом расширения содержания самого понятия «онтология» и расширения состава функций, приписываемых онтологиям, интересно также рассмотреть изменение определяемого онтологическим сообществом соотношения онтологических языков и различных традиционных языков информатики.

Собственно, онтологические языки относительно просты. Для представления элементов интенционального уровня нужны средства представления понятий, свойств понятий, связей между понятиями и их свойствами, аксиом. Аксиомы представляют собой логические формулы, которые выражают условия интенционального уровня, которым должны удовлетворять элементы экстенционалов. Представимы такие виды аксиом как выражение отношения «быть подклассом», отношения эквивалентности классов или их композиций, отношения непересечения классов, выражения условий (restrictions) кардинальности, и др.

На экстенциональном уровне представляются индивиды и факты. Индивид, или объект, представляется как экземпляр экстенционала понятия. Факты представляются как связи между экземплярами понятий или свойств.

Нетрудно видеть, что столь простые языки (или их подмножества) естественно отождествлять с подмножествами различных языков, используемых в тех или иных областях информатики. Собственно, такое отождествление нетрудно обнаружить в публикациях онтологического сообщества. Прежде всего, заметно стремление рассматривать онтологии как диаграммы (представляемые в виде семантических сетей, схем модели сущность – связь, диаграмм классов языка UML) [18].

Точка зрения онтологического сообщества на соотношение онтологических языков с другими языками выражается следующим образом [17]:

- онтологические языки по отношению к языкам представления знаний: онтологии являются схемами представления знаний;
- онтологические языки по отношению к логике: логика это инструмент для придания семантики онтологическим языкам;
- онтологические языки по отношению к концептуальным моделям данных: концептуальные схемы являются онтологиями специального вида, пригодными для концептуализации конкретной логической модели данных;

<sup>1</sup> В данном случае имеются ввиду схемы баз данных

<sup>2</sup> В контексте разговора об уровне абстракции данные часто превращаются в знания

<sup>3</sup> Т.Груббер намеренно забывает о концептуальных схемах баз данных

<sup>4</sup> Эта фраза удивительна: как будто не существуют не менее развитые (чем онтологические) языки для интеграции баз данных, концептуального моделирования, обеспечения семантической интероперабельности, спецификации посредников, позволяющие реализовать перечисленные функции

<sup>5</sup> В дедуктивных базах данных эти уровни давно принято называть интенциональной (IDB) и экстенциональной базами данных (EDB)

- онтологические языки по отношению к языкам программирования: определения классов в программах являются специальными онтологиями, служащими для концептуализации конкретных структур, используемых в вычислениях.

Эта схема отражает стремление позиционировать онтологические языки среди различных языков информатики, одновременно определяя для последних их место среди онтологических языков<sup>1</sup>. Часто в онтологическом сообществе к классу онтологических языков относят языки, которые никогда таковыми не были. Вот, например, одна из классификаций онтологических языков [17]:

1. Графовые:
  - Семантические сети;
  - Концептуальные графы;
  - Диаграммы классов UML, схемы модели сущность – связь;
2. Фреймовые:
  - Системы на фреймах;
  - ОКВС, XOL;
3. Логические:
  - Дескриптивные логики (например, SHOIQ, DLR, DL-Lite, OWL);
  - Языки правил (например, RuleML, LP/Prolog, F-Logic);
  - Логики первого порядка (KIF);
  - Неклассические логики (например, немонотонные, вероятностные логики).

Попытаемся разобраться, что нового дает такая экспансия онтологической терминологии. Может создаться впечатление, что онтологии открывают дорогу новым направлениям после десятилетий развития и исследований в области концептуальных моделей данных, дедуктивных баз данных, систем интеграции неоднородных баз данных, семантической интероперабельности, и пр. Но ведь в таких областях необходимый теоретический фундамент и конкретные высокоуровневые модели, языки давно созданы. В действительности, следовало бы сосредоточиться на выявлении полезности применения онтологических языков в различных областях информатики, в том числе для концептуального моделирования в области информационных систем и баз данных. Что же нового привносят в теорию и практику баз данных и информационных систем исследования в области онтологических языков?

Онтологические модели, рассматриваемые в публикациях, посвященных их использованию в базах данных и информационных системах (БД и ИС), основаны на логике предикатов первого порядка, чаще всего на ее подмножествах – дескриптивных логиках. По существу, в контексте БД и ИС онтологические языки (в частности, языки на дескриптивных логиках) играют роль моделей данных, и не более. Таким образом, следует сосредоточиться на анализе особенностей языков на дескриптивных логиках и новизны, привносимой ими в контекст БД и ИС (что могут дать модели данных на дескриптивных логиках в сравнении с реляционными, объектными и другими моделями данных).

Принципиально важным при этом является соотношение концептуального и онтологического моделирования. Подробное рассмотрение этого вопроса дано в [19]. Здесь достаточно будет ограничиться лишь несколькими основными положениями.

Концептуальное моделирование реализует абстрактное, семантическое моделирование предметной области (определение классов объектов предметной области, их взаимосвязей, ограничений), независящее от реализации, и служащее в качестве средства порождения эталонной спецификации, отражающей консенсус в сообществе, включающем разработчиков, пользователей ИС, и, собственно, самих ИС.

Концептуальные схемы применяются также в качестве глобальных схем при интеграции информационных ресурсов (баз данных), в процессе проектирования ИС и в процессе интерпретации запросов. Концептуальная схема определяет структуру предметной области, тогда как онтология должна быть ориентирована главным образом на определения используемых в предметной области понятий. Существенно, что концептуальные схемы БД и ИС, помимо описания классов объектов предметной области и ограничений, содержат описания поведения объектов (методов, функций, процессов), чего онтологии не содержат.

<sup>1</sup> Этот список нетрудно продолжить. Скажем, схемы баз данных – это онтологии специального вида для концептуализации структур баз данных, и т.д.

Вместе с тем, онтологическая модель позволяет определять в концептуальной схеме зависимости между классами объектов предметной области и их атрибутами, предполагающие осуществление логического вывода при ответе на запрос. Кроме того, онтологическими определениями понятий могут быть аннотированы соответствующие определения концептуальной схемы – вот где онтологии проявляют себя оригинальным образом

### 3 Развитие языков на дескриптивных логиках в контексте БД и ИС

В этом разделе дан краткий анализ развития языков на дескриптивных логиках, имеющих отношение к контексту БД и ИС. Этот анализ будет подразделен на две части. В первой рассматриваются языки на дескриптивных логиках, созданные до их включения в стек W3C. Во второй части основное внимание будет сосредоточено на соотношении языков на дескриптивных логиках (и других языков) в стеке W3C.

С конца 80-х годов 20 века проводились исследования и разработки в области дескриптивных логик, ориентированных на их использование в контексте баз данных. Одним из первых в этом ряду является язык CLASSIC [20], позволяющий определять структуры объектов базы данных не только в терминах их отношений с другими объектами, но также и интенционально. Язык обладает способностью вывода, например, определения классов, к которым данный объект принадлежит. Цели и результаты создания языка обсуждались еще в то время в терминах, близких к тем, что можно встретить сегодня в работах по «онтологически базированному» доступу к базам данных. Конечно, при этом не упоминается термин «онтология». Несколько позже стали развиваться языки, ориентированные на поддержку определений понятий предметных областей, на терминологическую интеграцию. К этому ряду систем можно отнести LOOM [21], GRAIL [22], SIMS [23], OIL [24]. Эти языки поддерживают практически все операции формирования понятий. LOOM обеспечивает также поддержку рассуждений над ABox и TBox.

Заметной вехой этого периода является появление языков, объединяющих возможности дескриптивных логик с программами на правилах. Язык AL-log [4,25] представляет собой комбинацию простой дескриптивной логики с языком Datalog. Язык CARIN [26] расширяет дескриптивную логику Хорновскими правилами без функций. Работы по интеграции дескриптивных логик с рассуждениями на правилах выявили необходимость ограничения выразительной способности терминологической части языка для сохранения разрешимости. Интересно, что уже в это время дескриптивные логики использовались в экспериментах по интеграции данных. Так, язык CARIN и дескриптивная логика DLR на n-арных отношениях использовались в работах по интеграции реляционных баз данных [26].

Этот период был отмечен также работами по развитым языкам представления информации, основанным на логике первого порядка, объектных и фреймовых моделях. Такими языками являются онтологический язык Ontolingua [27], язык СИНТЕЗ [28], ориентированный на обеспечение семантической интероперабельности и интеграции неоднородных информационных ресурсов, логико-фреймовый язык F-Logic [29], язык ОКВС [30] для обеспечения интероперабельности баз знаний. В сравнении с языками на дескриптивных логиках, языки на основе логики первого порядка и объектно-фреймовых моделей обладают большим разнообразием возможностей определения типов, классов, логических утверждений.

По мере развития деятельности W3C в области Семантического Веба все больше появляется работ в этом контексте по онтологическим языкам на дескриптивных логиках. Языки, признанные W3C в качестве технологической основы Семантического Веба, организуются в виде Стека Семантического Веба (рис. 1). Этот стек определяет иерархию языков, в которой каждый уровень опирается на средства нижележащих уровней. Он также демонстрирует

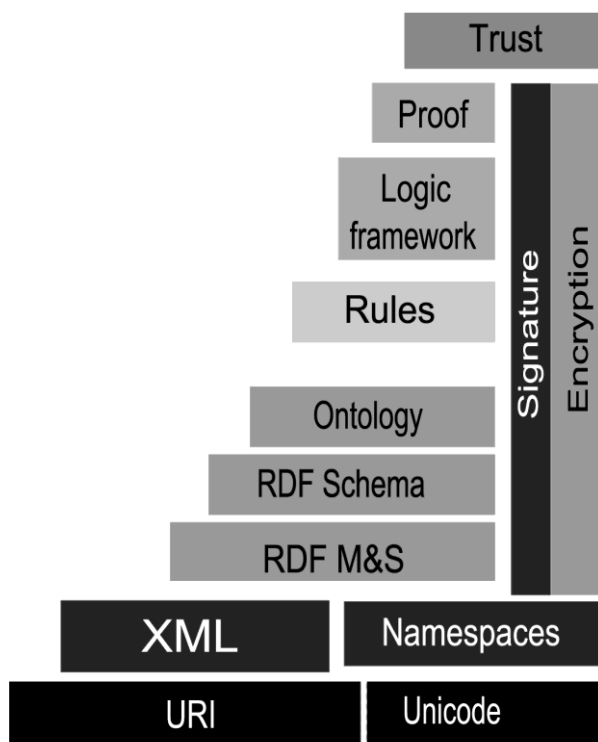


Рис. 1. Одностековая архитектура W3C

развитие Семантического Веба как расширения классической гипертекстовой Всемирной паутины. Основные языки Семантического Веба, имеющие непосредственное отношение к онтологическим спецификациям, включают: 1) RDF – простой язык для представления данных в виде триплетов, ссылающихся на объекты («ресурсы») Веба и определяющие их связи. Структуры данных, выразимые на RDF, представимы в синтаксисе XML; 2) язык RDF Schema, расширяющий RDF и позволяющий определять схемы свойств и классов ресурсов, представимых в RDF; 3) семейство языков OWL (OWL 2) для представления онтологий, имеющих формальную семантику, а также основанную на RDF/XML сериализацию таких онтологий.

В последнее время жизнеспособность одностековой архитектуры (SSA), развиваемой W3C, подвергается сомнению. Основное соображение заключается в том, что едва ли (особенно, принимая во внимание младенческий возраст Семантического Веба) единственная иерархия языков окажется достаточной для всех будущих семантических потребностей Веба [31]. Ведь любая технология, включая технологию создания языков, в конце концов устареет. Кроме того, никакая технология не может справиться со всеми проблемами. Более реалистичным представляется мультитековая архитектура (MSA), в которой множество стеков могло бы сосуществовать бок о бок друг с другом. Под влиянием языков правил, включенных в состав стека W3C в результате реализации проекта RIF (Rule Interchange Format), произошло раздвоение стека (рис. 1) и превращение его в мультистек MSA (рис. 2).

MSA является расширяемым, так что другие стеки могут быть добавлены к нему при необходимости. Каждый уровень стека в мультистеке является синтаксическим и семантическим расширением предыдущего уровня. В этом аспекте интересно посмотреть, что происходит в точке раздвоения стека. Языки правил являются языками логического программирования, исповедующими немонотонные рассуждения в отличие от языков OWL, основанных на дескриптивной логике. Характерным является (неявное) присутствие в MSA (рис. 2) уровня языка программ на дескриптивной логике (Description Logic Programs (DLP)), все операторы которого отображаются в Хорновские правила (подмножество логики первого порядка (рис. 3)) [32].

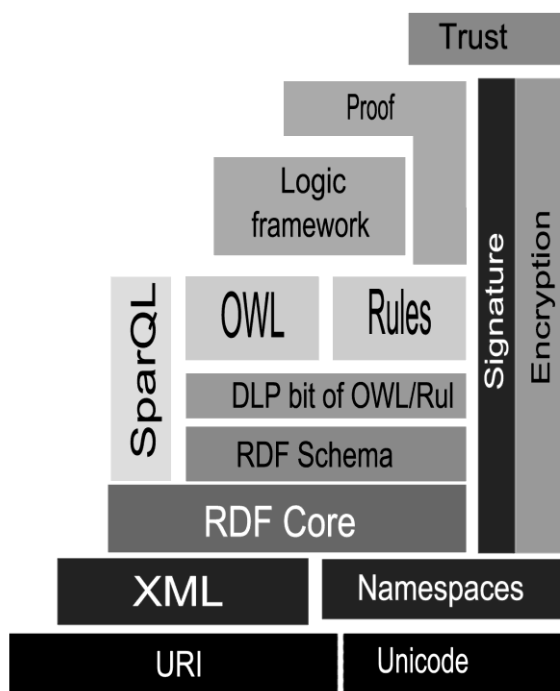


Рис. 2. Мультистековая архитектура W3C

Этот уровень должен хотя бы незримо присутствовать в стеке, для обеспечения совместимости в стеке снизу вверх: семантики DLP в стеке OWL и в стеке правил остаются одинаковыми. Вместе с тем, логические программы, основанные на подмножестве логики первого порядка (например, OWL), не поддерживают понятия ограничений целостности и их нарушений. Вместо этого используются зависимости как утверждения, определяющие желательное состояние среды. Например, если при выводе при условии моногамии оказывается, что у Ивана две жены – Анна и Вера, то OWL приходит к заключению, что Анна и

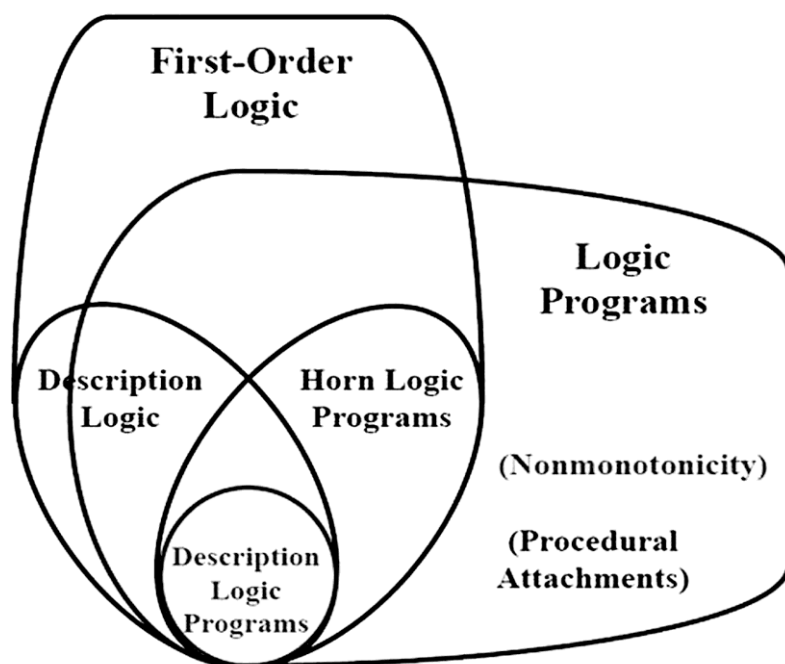


Рис. 3. Соотношение логических языков

Вера – это одно и то же лицо. В отличие от этого, программы на правилах с немонотонной моделью рассуждений будут трактовать такую ситуацию как противоречие в базе данных.

Более того, имеется мнение, согласно которому утверждение о том, что такие языки на правилах как Datalog являются расширением уровня DLP, неверно. Например, если в базе данных имеется единственный факт *знает(иван, анна)*, то DLP и Datalog дадут разные ответы на запрос, знает ли *иван* в точности одно лицо. Согласно семантике OWL ответ будет “*unknown*”, тогда как согласно семантике языков на правилах ответом будет “*yes*”. Однако в действительности оба ответа верны. Все зависит от того, какой стек выбирает пользователь – стек правил, или стек OWL.

Вместе с тем создание MSA потребует еще значительных усилий. Например, открытым остается вопрос о месте профилей языка OWL 2, связанных с реляционной моделью данных, учитывая противоречия между ограничениями и зависимостями, о которых речь шла выше. Интересно, что создание профиля OWL 2 QL [33] оказалось возможным благодаря исследованиям, проведенным по семейству дескриптивных логик DL-Lite [3,34].

Целью создания DL-Lite является поддержка базовых средств онтологических языков при сохранении приемлемой сложности рассуждений (в функции рассуждений включаются также ответы на запрос, представляемый как объединение конъюнктивных запросов над экстенциональным уровнем (ABox)). Сложность рассуждений DL-Lite является полиномиальной по размеру TBox, а сложность ответа на запрос оценивается как  $AC_0$  от размера ABox.

Существенно, что DL-Lite во время обработки запросов позволяет вести рассуждения в TBox и ABox независимо, так что вначале реализуются рассуждений над TBox, а затем запрос над ABox может быть реализован с помощью SQL.

Логика, включенные в семейство DL-Lite, являются *максимальными дескриптивными логиками*, обеспечивающими эффективный ответ на запрос над базами данных большого объема.

Семейство DL-Lite образуется следующим образом (рис. 4).

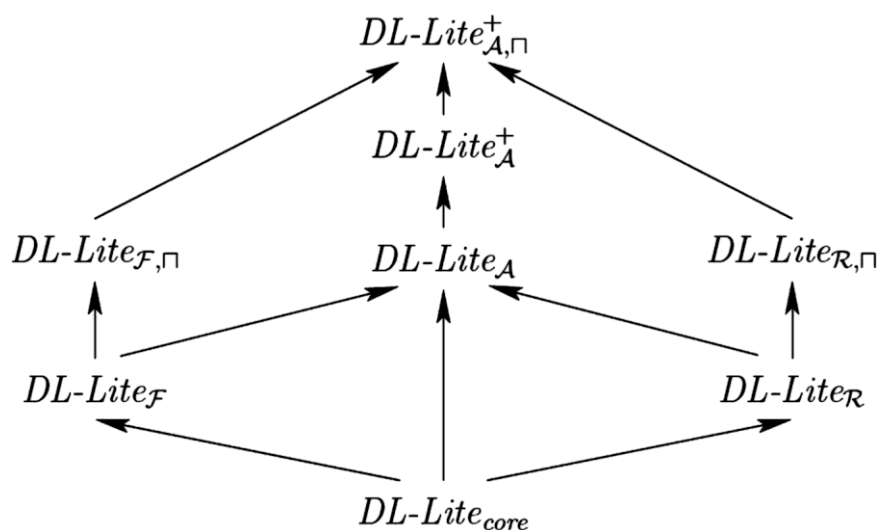


Рис. 4. Семейство дескриптивных логик DL-Lite

DL-Lite<sub>core</sub> поддерживает утверждения ISA над понятиями, условия непересечения понятий, типизацию ролей, ограничения включения. В DL-Lite<sub>F</sub> добавляются средства выражения условий функциональной зависимости на ролях. DL-Lite<sub>R</sub> расширяется утверждениями ISA над ролями, условиями непересечения ролей. В DL-Lite<sub>A</sub> добавляются возможности совместного использования утверждений включения ролей и утверждений функциональности.

Профиль OWL 2 QL основан на DL-Lite<sub>R</sub> [33]. Существенно, что этот профиль включает около трети набора аксиом, поддерживаемых OWL 2 [33]. Вместе с тем, простые дескриптивные логики, подобные DL-Lite<sub>R</sub>, все же годятся для поддержки простых онтологических языков и концептуальных моделей данных (подобных модели Сущность – Связь и подмножеству диаграмм классов UML).



## 4 Онтологически базированный доступ к данным

«Онтологически базированные» технологии предполагают использование информационных ресурсов на основе концептуализации предметных областей и доступ к данным, опосредованный онтологией (концептуальным взглядом на данные). Собственно, исследовательских работ в этой области известно немного. Применяемые в них подходы рассматриваются в настоящем разделе. Следует сразу заметить, что эти работы применимы только к реляционным базам данных. Собственно, для обеспечения доступа к данным при использовании онтологий требуется 1) определить отображение схем реляционных баз данных в онтологическую (концептуальную) схему; 2) преобразовать конъюнктивный запрос в терминах концептуальной схемы в запрос над реляционной базой данных; 3) реализовать полученный запрос. Рассмотрим последовательно, как это делается. Рассматриваемый подход вполне приемлем для реализации профиля OWL 2 QL.

Реляционные базы хранят данные, в то время как экземплярами понятий (классов) являются объекты. Основной механизм преодоления такого несоответствия импеданса хорошо известен из области объектных баз данных. При этом уникальный идентификатор объекта продуцируется Сколемовской функцией, параметром которой является значение кортежа базы данных. Отображение реляционной модели данных в концептуальную конструируется как совокупность конъюнктивных запросов над атомарными понятиями, атрибутами, ролевыми атрибутами и соответствующих им запросов на языке SQL. Формально, отображение состоит из утверждений вида  $\varphi \rightsquigarrow \psi$ , где  $\varphi$  – это произвольный SQL запрос арности  $n > 0$  над базой данных, а  $\psi$  – объединение конъюнктивных запросов арности  $n > 0$  в терминах онтологии  $T$  (например, TBox'a в DL-Lite<sub>R</sub>).

Для ответа на запрос каждое утверждение отображения  $\varphi \rightsquigarrow \psi$  расщепляется на несколько утверждений вида  $\varphi \rightsquigarrow p$ , по одному для каждого атома  $p$  в  $\psi$ . Атомы запроса  $q$  унифицируются с атомом правой части отображения, в результате чего образуется объединение конъюнктивных запросов. Затем каждый атом разворачивается, используя левую часть отображения, для получения SQL запроса.

Следует заметить, что перед выполнением описанной выше процедуры запрос дедуктивно расширяется на основе аксиом, заданных в схеме, и машины вывода DL-Lite<sub>R</sub>. Кроме того, согласно онтологической модели, информация о предметной области неполна. Полученные в результате преобразования запросы обеспечивают получение *достоверных* (certain) ответов, независимо от того, каким образом может быть произведено пополнение неполной информации.

Описанный подход реализован в QuOnto [35] – инструментальном средстве, поддерживающем рассуждения в онтологиях семействах DL-Lite, а также ответы на запросы, представленные в виде UCQ (объединения конъюнктивных запросов). QuOnto допускает подключение внешних реляционных СУБД (таких как Oracle, DB2, IBM Information Integrator, SQL Server, MySQL, и др.).

Вместо реляционной СУБД имеется возможность совместной работы QuOnto с федеративной базой данных – результатом интеграции реляционных баз данных. Такая компоновка QuOnto именуется MASTRO-I [36]. Нужно иметь в виду, что федеративная база данных представлена для QuOnto единственной схемой отношения. Для формирования федеративной базы данных используется подход GAV. Функционирование QuOnto в MASTRO-I ничем не отличается от уже описанного в настоящем разделе.

## 5. Заключение

В статье дан краткий анализ состояния работ в области «онтологически базированных» систем доступа к данным и их возможного влияния на развитие информационных систем и баз данных. Обсуждены вопросы соотношения онтологического и концептуального моделирования и соответствующих языковых средств. Дан анализ развития языков на дескриптивных логиках, ориентированных на использование в качестве средств концептуального моделирования в

контексте баз данных и информационных систем. Приведен краткий обзор экспериментальных реализаций «онтологически базированных» систем доступа к данным.

Развитие дескриптивных логик, ориентированных на применение в контексте баз данных и информационных систем, и известные экспериментальные результаты создания «онтологически базированных» систем доступа к данным показывают, что заметным результатом является создание семейства дескриптивных логик DL-Lite как максимальных подмножеств средств, обладающих приемлемой для работы с базами данных эффективностью. Эти результаты нашли применение в профиле OWL 2 QL [33], базирующемся на DL-Lite<sub>R</sub>. Работы по «онтологически базированным» информационным системам ориентированы на применение онтологических языков для концептуального моделирования, т.е. на создание концептуальных моделей данных на дескриптивных логиках. Языки на дескриптивных логиках и соответствующие им концептуальные модели могут быть отображены в существующие модели данных систем интеграции баз данных (например, в язык СИНТЕЗ [28]) с сохранением их семантики [37]. Более того, базы данных со схемами в OWL могут быть интегрированы с другими, традиционными, базами данных в системах интеграции баз данных. Отображение OWL в СИНТЕЗ, сохраняющее семантику, построено [37,38].

Представляется, что с точки зрения баз данных и информационных систем идея так называемых «онтологически базированных» систем является скорее отражением терминологической экспансии онтологий в область баз данных. Вместе с тем, выполненные в этой сфере исследования и разработки в области дескриптивных логик важны для включения реляционных баз данных в контекст Семантического Веба.

Альтернативный подход к эффективной поддержке запросов над схемами баз данных с онтологическими зависимостями, развиваемый Георгом Готтлобом и его коллегами в традициях технологий баз данных [5-12], представляется весьма перспективным. Подход заключается в выделении максимальных фрагментов языка логики первого порядка, позволяющих выражать онтологические зависимости в схемах баз данных и сохраняющих возможность эффективной поддержки запросов к базам данных над такими схемами. Определено семейство языков Datalog $\pm$ , обладающее свойствами простоты восприятия, разрешимости и эффективности. В области онтологических рассуждений полученные языки являются чрезвычайно гибкими и выразительными. Доказано, что простейший из языков семейства Datalog $\pm$  (линейный Datalog $\pm$  вместе с зависимостью отрицания и неконфликтующими ключами) обладает выразительной способностью, превышающей возможности логик семейства DL-Lite. Анализ семейства языков Datalog $\pm$  и их сопоставление с дескриптивными логиками семейства DL-Lite предполагается дать в отдельной статье.

## Литература

1. Diego Calvanese<sup>1</sup>, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Riccardo Rosati. Ontology-based database access. // Proc. of the 15th Italian Conf. on Database Systems (SEBD 2007), 2007.
2. Diego Calvanese. Ontology-based Data Management Masters Ontology Spring School, September , 2009, // <http://ksg-projects.meraka.csir.co.za/krr-projects/events/moss09-1/MOSS-09-OBDM-calvanese-draft.pdf>.
3. Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Mariano Rodriguez-Muro<sup>1</sup>, and Riccardo Rosati Reasoning Ontologies and Databases: The DL-Lite Approach Web 2009, // LNCS 5689, Springer-Verlag, 2009, pp. 255–356.
4. F. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. Al-log: Integrating datalog and description logics. // Journal of Intelligent Information Systems (JIIS), 27(1), 1998.
5. Andrea Cali, Georg Gottlob, Thomas Lukasiewicz. Datalog $\pm$  : A Unified Approach to Ontologies and Integrity Constraints. // ICDT 2009, March, 2009, Saint Petersburg, Russia.
6. Andrea Cali, Georg Gottlob, Thomas Lukasiewicz. A General Datalog-Based Framework for Tractable Query Answering over Ontologies. // PODS'09, June, 2009.
7. Andrea Cal<sup>1</sup>, Georg Gottlob, Andreas Pieris. Advanced Processing for Ontological Queries.

- // Proceedings of the 36th International Conference on Very Large Data Bases, September, 2010, Singapore.
8. Andrea Cali, Georg Gottlob, Thomas Lukasiewicz, Bruno Marnette, Andreas Pieris. Datalog+/-: A Family of Logical Knowledge Representation and Query Languages for New Applications. // Proceedings of the 25th Annual IEEE Symposium on Logic in Computer Science, 2010, pp.228-242.
  9. Andrea Cali, Georg Gottlob, and Andreas Pieris. Query Answering under Expressive Entity-Relationship Schemata. // ER 2010, LNCS 6412, , 2010, pp. 347–361.
  10. Andrea Cali, Georg Gottlob and Andreas Pieris. New Expressive Languages for Ontological Query Answering. // Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011
  11. G.Gottlob. Ontological Queries Rewriting and Optimization, // Proceedings of the ICDE, 2011.
  12. Georg Gottlob, Giorgio Orsi, and Andreas Pieris. Ontological Query Answering via Rewriting. // ADBIS 2011, LNCS 6909, 2011, pp. 1–18.
  13. Л. А. Калиниченко. Экспансия онтологий: онтологии в информационных системах. // Труды Второго Симпозиума «Онтологическое моделирование», Казань, 2010, М.: ИПИ РАН, 2011. – С. 29 – 44.
  14. Thomas R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. // International Journal Human-Computer Studies 43, 1995, p. 907-928.
  15. Nicola Guarino Formal Ontology in Information Systems. // Proceedings of FOIS'98, Trento, Italy, June 1998. Amsterdam, IOS Press, pp. 3-15.
  16. T. Gruber. "Ontology". // *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2008.
  17. Diego Calvanese. Knowledge Bases and Databases. Part 2: Ontology-Based Access to Information. // Presentation, 2009, <http://www.inf.unibz.it/~calvanese/teaching/08-09-kbdb/lecture-notes/p2-obda-2up.pdf>.
  18. Diego Calvanese. Ontology-based Data Management. // Masters Ontology Spring School, September , 2009, <http://ksg-projects.meraka.csir.co.za/krr-projects/events/moss09-1/MOSS-09-OBDM-calvanese-draft.pdf>.
  19. Kogalovsky M.R., Kalinichenko L.A. Conceptual and Ontological Modeling in Information Systems. // Programming and Computer Software. -- Moscow: MAIK Nauka/Interperiodica, Pleiades Publishing Inc., 2009. -- V. 35, N. 5. -- P. 241–256.
  20. Borgida, R. J. Brachman, D. L. McGuinness, L.A. Resnick. CLASSIC: A Structural Data Model for Objects, // ACM SIGMOD Record, Volume 18, Issue 2, June 1989.
  21. Robert M. MacGregor. Using a description classifier to enhance deductive inference. // Proceedings of the Seventh IEEE Conference on AI Applications, pp. 141–147, 1991.
  22. A.L. Rector, S. Bechofer, C.A. Goble, I. Horrocks, W.A. Nowlan, and W.D. Solomon. The grail concept modelling language for medical terminology. // Artificial Intelligence in Medicine, 9:139 – 171, 1997.
  23. Yigal Arens, Chun-Nan Hsu, and Craig A. Knoblock. Query processing in the sims information mediator. In Advanced Planning Technology. // AAAI Press, California, USA, 1996.
  24. D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, and M. Klein. Oil in a nutshell. // 12th International Conference on Knowledge Engineering and Knowledge Management EKAW2000, France, 2000.
  25. Diego Calvanese, Giuseppe De Giacomo, and Maurizio Lenzerini. Description logics for information integration. // Computational Logic: From Logic Programming into the Future, LNCS, Springer-Verlag, 2001.
  26. Francois Goasdou'e, V'eronique Lattes, and Marie-Christine Rousset. The use of carin language and algorithms for information integration: The picsele project. // International Journal of Cooperative Information Systems (IJCIS), 9(4):383 – 401, 1999.
  27. Tom Gruber. A translation approach to portable ontology specifications. // Knowledge Acquisition, 5(2):199–220, 1993
  28. Калиниченко Л.А. СИНТЕЗ – язык определения, проектирования и программирования сред неоднородных интероперабельных информационных ресурсов. // ИПИ РАН, 1993, 110 стр.

29. M. Kifer, G. Lausen, and J. Wu. Logical foundations of object-oriented and frame-based systems. // *Journal of the ACM*, Volume 42 Issue 4, July, 1995.
30. Vinay K. Chaudhri, Adam Farquhar, Richard Fikes, Peter D. Karp, and James P. Rice. Open knowledge base connectivity (okbc) specification document 2.0.3. // Technical report, SRI International and Stanford University (KSL), April 1998.
31. Michael Kifer, Jos de Bruijn, Harold Boley, and Dieter Fensel. A realistic architecture for the semantic web. // *Proceedings of the 1st International Conference on Rules and Rule Markup Languages for the Semantic Web (RuleML2005)*, LNCS, 3791, Galway, Ireland, November 2005. Springer, pp. 17–29.
32. Benjamin N. Grosz, Ian Horrocks, Raphael Volz, Stefan Decker. Description Logic Programs: Combining Logic Programs with Description Logic. // *WWW2003*, May 20–24, 2003, Budapest, Hungary.
33. OWL 2 Web Ontology Language: Profiles, // W3C, 2009.
34. Diego Calvanese Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Riccardo Rosati. Tractable Reasoning and Efficient Query Answering in Description Logics: The DL-Lite Family. // *Journal of Automated Reasoning*, Volume 39, 2007.
35. Acciarri, D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, M. Palmieri, and R. Rosati. QUONTO: QUerying ONTOlogies. // *Proc. of AAAI 2005*, pp. 1670–1671.
36. Diego Calvanese, Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi, Riccardo Rosati. MASTRO-I: Efficient integration of relational data through DL ontologies. // *Proceedings of the 2007 International Workshop on Description Logic (DL 2007)*, CEUR Electronic Workshop Proceedings, 2007.
37. Leonid Kalinichenko, Sergey Stupnikov. OWL as Yet Another Data Model to be Integrated. *Proceedings of the 15th International Conference*, // *ADBIS 2011*, Vienna, Austria, September, 2011.
38. С. А. Ступников, Н. А. Скворцов. Взаимное отображение канонической информационной модели и языка OWL 2. // *Труды 12-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»*, RCDL'2010. – Казань: Казанский государственный университет, 2010. – С. 392-398.

## **Support of databases with ontological dependencies on the basis of description logics**

L.A.Kalinichenko

leonidk@synth.ipi.ac.ru

Institute of Informatics Problems of RAS

**Abstract** The paper provides brief analysis of the state of the art in the area of “ontology based” data access systems and of the possible influence of such works on the development of research in the database and information systems fields. The issues of correlation between ontological and conceptual modeling and between the respective languages are discussed. An analysis of development of the languages based on description logics and oriented on their use as a facility for conceptual modeling in the database context is given in brief. Short overview of experimental results of development of “ontology based” data access systems is also provided.

**Keywords** ontologies, ontological data dependencies, conceptual modeling, description logics, relational data models, Datalog, information systems.