



НОВІ ЗАСОБИ КІБЕРНЕТИКИ, ІНФОРМАТИКИ, ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ ТА СИСТЕМНОГО АНАЛІЗУ

Ф.И. АНДОН, В.А. РЕЗНИЧЕНКО, И.С. ЧИСТЯКОВА

УДК 004.62

ОТОБРАЖЕНИЕ ДЕСКРИПТИВНОЙ ЛОГИКИ В РЕЛЯЦИОННУЮ МОДЕЛЬ ДАННЫХ

Аннотация. Обсуждены общие задачи проблемы интеграции данных, а также онтолого-ориентированная интеграция данных в семантическом вебе. Показана важность установления отображений между онтологиями и реляционными базами данных, дан краткий аналитический обзор исследований по этой теме. Определена бинарная реляционная модель данных и описано отображение дескриптивной логики ALC и ее расширений в данную модель.

Ключевые слова: интеграция данных, онтология, семантический веб, дескриптивная логика, бинарная реляционная модель данных.

ВВЕДЕНИЕ. ОПРЕДЕЛЕНИЕ ПРОБЛЕМЫ ИНТЕГРАЦИИ ДАННЫХ

Интеграция данных — одно из наиболее востребованных направлений в современной информационной индустрии. Однако оно не ново. Первые шаги в этой области сделаны в середине 70-х годов. В это время начинали формироваться более четкие представления о многоуровневой архитектуре баз данных (БД), о моделях данных (МД) как инструменте информационного моделирования реальности и об отображении между ними. При этом акцентировалось внимание, главным образом, на поддержке глобальной схемы, получившей название концептуальная, для совокупности локальных схем, поддерживающих одну и ту же (или различные) МД, функционирующих в различных узлах под управлением одной СУБД. Впоследствии задача была обобщена и включала создание хранилищ данных, различных репозиторий информационных ресурсов и веб-приложений. В настоящее время ключевой в интеграции данных является интеграция неоднородных (гетерогенных) данных различных информационных систем.

За годы существования интернет-пространства (особенно в последнее десятилетие) в нем скопилось большое количество информации, объем которой возрастает в геометрической прогрессии. Это порождает множество актуальных проблем, связанных с использованием и хранением данных информационного пространства. Огромные объемы данных в гетерогенных источниках представляют информацию различными способами и имеют разнообразное функциональное назначение. Интеграция и совместное использование информации из множества таких источников данных — сложная задача, неизменно актуальная на протяжении последних десятилетий.

Рассмотрим несколько причин возникновения гетерогенности.

Различные модели данных. Известно, что от 75 до 90% информации хранится в реляционных БД (РБД). Однако на остальные 10–25% приходится немалое количество данных, хранящихся в структурах, которые определяются совершенно другими МД со своей специфической семантикой.

Различные способы хранения данных (файлы, БД, хранилища и т.д.). Физическая организация хранения информации создает дополнительные препятствия

© Ф.И. Андон, В.А. Резниченко, И.С. Чистякова, 2017

для ее использования. Интеграция данных должна предоставить единый логический формат организации данных так, чтобы независимо от способа их физического хранения конечный пользователь имел единый механизм доступа к содержимому.

Существенная изолированность данных. Источники информации изолированы один от другого и каждый из них подчиняется концепции «замкнутого мира». Такой подход значительно затрудняет введение принципиально новых понятий различных предметных областей, порождает дублирование данных, что приводит к увеличению объема, но уменьшению релевантности искомой информации. Интеграция данных способствует устранению такой изоляции источников и согласованному использованию уже существующих данных.

Противоречивость данных. Изоляция источников приводит к возникновению противоречивости. Интеграция данных призвана устранить этот недостаток путем введения единого семантического контекста для всех информационных ресурсов, хранящихся в интегрированных источниках.

Различные способы оперирования данными (манипулирование, поиск, выборка и т.д.). Существующие возможности поисковых систем общего назначения не позволяют обеспечить универсального поиска информации. Каждая МД предполагает существование собственных средств манипулирования, что порождает их разнообразие, приводящее к гетерогенности данных.

Ввиду перечисленного очевидна важность решения комплексной проблемы интеграции.

Проблема интеграции данных. Она заключается в таком логическом объединении данных, принадлежащих разнородным источникам, которое обеспечивает единое представление и оперирование ими. Система интеграции данных позволяет освободить пользователя от необходимости самостоятельно выбирать источники, в которых находится интересующая пользователя информация, обращаться к каждому источнику по отдельности и вручную сопоставлять и объединять данные из различных источников.

Согласно [1] проблема интеграции комплексная и многоаспектная. В то время, как ее основной целью является обеспечение унифицированного представления и манипулирования данными различных источников, конкретная задача интеграции может зависеть от множества разнообразных прикладных факторов.

1. СОСТАВЛЯЮЩИЕ ПРОБЛЕМЫ ИНТЕГРАЦИИ ДАННЫХ

Выделяют три основных составляющих проблемы интеграции данных: схемы их интеграции, отображения между МД, способы манипулирования данными.

Рассмотрим их кратко, с более детальным обсуждением можно ознакомиться в [2].

Схемы интеграции данных. Существуют две основные схемы интеграции данных: P2P (peer-to-peer), ее также называют одноранговой, и централизованная. В P2P-схеме не существует глобальных точек контроля и управления. В основе каждого узла, имеющегося в схеме, лежит своя МД. Каждый узел этой схемы должен иметь возможность оперировать данными тех моделей, которые лежат в основе его узлов-соседей. Для решения комплексной проблемы интеграции данных наиболее успешной в настоящее время является централизованная схема. Эту архитектуру также называют архитектурой с посредником [1]. На ней основаны многие подходы к выработке отображений между МД системы и к разработке способов манипулирования данными.

В централизованной схеме (рис. 1) обычно имеется одна глобальная точка контроля (посредник). В основе этого узла лежит своя МД, в [3] она называется глобальной схемой, а все остальные модели — локальными, или схемами источников. Далее использована эта терминология. Основная роль глобальной схемы — предоставление пользователю единого интерфейса для доступа к информации, хранящейся в реальных источниках данных, а также поддержка отображений между глобальным и локальным представлениями данных. Глобальная схема обеспечивает единое согласованное представление данных и предоставле-

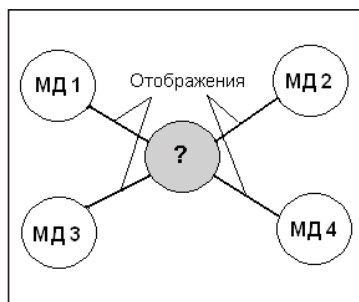


Рис. 1. Централизованная схема интеграции данных

ние централизованного поиска. Критическим моментом централизованной схемы является разработка отображений между МД источников и МД глобальной схемы.

Отображения между моделями данных.

Существует две разновидности централизованной архитектуры: Global-as-View (GAV) и Local-as-View (LAV) [3]. Дадим краткую характеристику каждой из них. Система интеграции данных (СИД) представляется тройкой $\langle G, S, M \rangle$, где G — глобальная схема, S — схема источника, M — отображения между G и S , имеется одна глобальная схема и множество схем

источников. Все схемы задаются в терминах собственных МД. Отображения M между G и S представляют собой набор утверждений в виде $q_S \rightarrow q_G$ и $q_G \rightarrow q_S$, где q_G и q_S — совместимые запросы, сформулированные в терминах моделей G - и S -схем соответственно. Запись $q_S \rightarrow q_G$ означает, что каждое понятие источника, представленное запросом q_S , соответствует понятию глобальной схемы, представленному запросом q_G , и наоборот.

В СИД $I = \langle G, S, M \rangle$, основанной на LAV-подходе, отображение M связывает каждый элемент s схемы источника S с запросом q_G схемы G . Таким образом, LAV-отображение — это набор утверждений, по одному на каждый элемент s из схемы S , в форме $s \rightarrow q_G$. С точки зрения моделирования подход LAV основан на следующей идее: каждый элемент источника s должен быть связан запросом q_G с соответствующим элементом глобальной схемы. Добавление нового источника сводится к обогащению набора отображений новыми утверждениями без прочих изменений.

В GAV-подходе отображения M связывают каждый элемент g схемы данных G запросом q_S источника S . Таким образом, GAV-отображение — это набор утверждений, по одному на каждый элемент g из схемы G , в форме $g \rightarrow q_S$. Отображение устанавливает, как нужно извлечь данные из источника, когда следует оценить различные данные глобальной схемы.

Каждый из этих подходов имеет преимущества и недостатки, но принципиально важно в них то, что отображением является не простое установление соответствия между структурными элементами МД, а запрос. Это означает, что в основе отображения между моделями лежит некоторый логический аппарат конкретного языка формулировки запросов.

В работе [4] предлагается создавать так называемые «обертки» для каждого источника. Они являются локальными схемами, представленными в той же МД, что и глобальная схема. Предполагается, что каждый информационный источник «обернут» промежуточным компонентом-адаптером, который отвечает за выборку сведений из источника в рамках единой МД, а также за предоставление стандартного технического интерфейса для обращения к источнику.

Способы манипулирования данными. После описания отображений между МД возникает вопрос об использовании таких систем интеграции данных, т.е. о способе манипулирования созданными глобальными схемами и управления данными, расположенными внутри различных источников. Так, например, если в основу интеграции данных положен онтологический подход, о чем пойдет речь далее, то основой манипулирования данными становятся операции манипулирования онтологиями (алгебра онтологий) [5].

2. ОНТОЛОГО-ОРИЕНТИРОВАННАЯ ИНТЕГРАЦИЯ ДАННЫХ

В [6] сформулировано понятие семантической интеграции данных как процесса использования концептуального представления данных, а также их взаимоотношений для ликвидации возможных неоднородностей.

Уточним это определение следующим образом. Семантическая (онтолого-ориентированная) интеграция данных — это использование онтологии в качестве объединяющей (глобальной) модели для описания и поддержания отображений между различными МД, а также унифицированного манипулирования данными.

Использование онтологий для семантической интеграции данных аргументируется такими факторами: онтологии базируются на самой развитой МД, имеют более развитую семантику, предоставляют мощные механизмы вывода, а также имеют формальную спецификацию (дескриптивная логика).

Далее понятие онтологии рассматривается в ее стандартном, классическом определении, которое было сформулировано много лет назад, а именно онтология — это формальная, явная спецификация согласованной концептуализации [7].

В случае использования централизованной схемы интеграции данных для решения комплексной проблемы интеграции онтология наилучшим образом подходит для глобальной схемы, что позволяет в качестве локальной схемы применять любую МД. Вопрос взаимодействия внутри такой системы интеграции относится к общей проблеме интеграции — выработке отображений между моделями.

Для строгого описания отображений между двумя моделями необходимо, чтобы обе имели формальную спецификацию. В связи с этим в качестве модели глобального уровня выбрана дескриптивная логика (ДЛ) — семейство языков представления знаний о прикладной предметной области в структурированном и формальном, хорошо понятном виде [8]. Наиболее известные языки описания онтологий, OWL и OWL2 включительно, основаны на ДЛ.

3. ОТОБРАЖЕНИЕ МЕЖДУ ОНТОЛОГИЯМИ И РЕЛЯЦИОННЫМИ БАЗАМИ ДАННЫХ

В централизованной схеме интеграции источниками могут являться различные МД. Одна из них, несомненно, — реляционная модель данных (РМД). Проблема установления соответствий между онтологиями и РБД, кроме научного, имеет большой практический интерес. Он заключается в следующих фактах:

— онтологии должны иметь физическое место расположения. Их можно хранить в виде обычных файлов — это наиболее простой вариант, однако с множеством недостатков (не обеспечивается масштабируемость, нет механизмов поддержки совместного использования и др.). Существует иной подход — использование для этих целей СУБД. В данном контексте его основной недостаток — это поддержка гипотезы замкнутого мира, в то время как онтологии используют гипотезу открытого мира. Тем не менее, в этом случае можно воспользоваться опытом хранения, манипулирования и управления данными, приобретенными за последние 50 лет;

— согласно [9] приблизительно 77,3% данных существующего веба содержится в РБД, из которых формируется огромное количество веб-страниц, что побуждает использовать накопленные за эти годы массивы данных для создания онтологий в семантическом вебе;

— последнее десятилетие стремительно развивается новая IT-отрасль — мобильные технологии. Появление принципиально новых устройств значительно изменило мировое общественное сознание и позволило рядовому обывателю интуитивно почувствовать все преимущества семантических данных, что подстегнуло переломное научное сообщество к дальнейшему интенсивному развитию и внедрению в жизнь семантических технологий. Однако в действительности 100% мобильных устройств используют для своей работы БД (или их адаптированную версию).

На основании приведенных фактов можно сделать вывод о том, что в контексте проблемы отображений первоочередной задачей является создание механизмов взаимодействия именно между РМД и онтологиями.

Приведем краткий обзор исследований по установлению отображений между онтологической и реляционной МД. (Более детальный обзор см. в [10].) Существует следующая классификация направлений исследований и разработок в области описания отображений между онтологической и реляционной МД [11, 12]:

- использование промежуточной МД;
- непосредственное отображение онтологии в РБД;
- непосредственное отображение РБД в онтологию;
- создание и использование онтолого-ориентированной БД.

Рассмотрим эти подходы.

3.1. Использование промежуточной МД. Данный подход подробно изложен в фундаментальной работе [3]. Его идея заключается в том, что преобразование проводится не непосредственно, а с использованием промежуточной модели, и предполагается, что описание отображений между такой моделью-посредником, с одной стороны, и онтологией и реляционной МД, с другой стороны, осуществляется намного легче и естественнее. Так, например, в [13] в качестве такого посредника предлагается использовать UML. В других исследованиях в качестве модели-посредника рассматривается ER. В связи с тем, что отображение ER–РМД давно исчерпывающе исследовано, то до последнего времени проблемным оставалось отображение онтология–ER. В связи с этим в [14] предложен подход по автоматическому отображению ER в OWL и, наоборот, с использованием ДЛ. В [15] также предложен метод автоматического отображения расширенной ER-модели в онтологию OWL.

3.2. Непосредственное отображение онтологии в БД. Суть этого подхода заключается в том, что предлагается алгоритм преобразования онтологии в БД. Практически все исследования по отображению онтология–база данных базируются на реляционной МД. Например, в [11] предлагается алгоритм приведения онтологий OWL в схему РБД, который преобразует классы, свойства, атрибуты и ограничения OWL в конструкции схемы РБД. Алгоритм преобразует OWL Lite и частично OWL в DL. В [16] также предлагается набор правил отображения OWL в SQL-базу данных. Предложенные 12 правил не только отображают такие структурные элементы, как классы и свойства, но также учитывают характеристики свойств и другие характеристики. В работе [17] предлагается отображать онтологию не непосредственно в РБД, а в концептуальную МД базы данных.

3.3. Непосредственное отображение БД в онтологию. Это наиболее популярное направление среди всех, имеющих отношение к установлению взаимосвязи между онтологиями и РБД. Выделяются следующие аспекты исследований и разработок в этом направлении: отображение схемы (метаданных) РБД в онтологию; отображение с учетом анализа данных БД; отображение с учетом анализа запросов к БД. Рассмотрим эти направления.

Отображение схемы (метаданных) RDB в онтологию. В [18] предлагается отображать таблицы в OWL-классы и затем обогащать их с использованием других информационных источников. В [19] предлагаются правила анализа и отображения конструкций реляционной МД (отношений, атрибутов, кортежей и ограничений целостности) в соответствующие семантически эквивалентные конструкции онтологии (классы, атрибуты, экземпляры, аксиомы). Различные аспекты установления такого отображения и описание соответствующих трансформирующих правил также приведены в [20–27].

Отображение с учетом анализа данных БД. Анализ БД может предоставить дополнительную информацию для решения проблемы отображения РМД–онтология. Например, если установлено, что домен атрибута одного отношения является поддоменом атрибута другого отношения, то это может стать основанием для заключения, что между этими отношениями, возможно, существует таксономическая иерархия. В [23] алгоритм отображения содержит не только правила, использующие схему БД, но и правила, которые учитывают возможные ограничения на данные БД (которые в схеме не задаются). В [18] предлагается подход, когда онтологии генерируются из «традиционных» таблиц (например, из полученных при публикации БД или ее фрагментов в HTML-страницах). Этот подход назван TANGO (Table ANalysis for Generating Ontologies) и предполагает распознавание структуры таблицы и извлечение из нее концептов; выявление и

анализ ограничений, существующих между концептами; сопоставление концептов с другими, более общими концептами; слияние результирующей структуры с другими аналогичными структурами представления знаний.

Отметим еще одно исследование в этой области [19]. Оно основано на идее использования семантики, полученной из анализа HTML-страниц с содержимым БД, для реструктурирования и обогащения схемы БД в целях ее последующего применения при построении онтологии OWL. Вопрос использования HTML-контента для более глубокого анализа содержимого БД также рассмотрен в [24].

Отображение с учетом анализа запросов к БД. Не только содержимое БД, но и запросы к ней могут стать источником дополнительной информации. Так, в [28] предлагается следующий метод. Вначале на основании анализа схемы и БД строится онтология. Затем осуществляется анализ запросов пользователей, извлекается из них дополнительная информация и на ее основании совершенствуется онтология. Как следует из примеров, в результате анализа запросов в онтологии могут появляться новые классы, свойства, таксономические отношения между классами и т.д.

3.4. Создание и использование онтолого-ориентированной БД. Проводятся исследования и разработки еще одного направления в области интеграции онтологий и БД — создание специализированных РБД для хранения и манипулирования онтологиями. Такие БД названы онтолого-ориентированными (Ontology Based DataBases, OBDB). Согласно [29] OBDB — это такая БД, в которой онтология и ее данные хранятся в единой общей модели. В идеале в OBDB [30] должны явно представляться онтология, схема данных, непосредственно данные, их связи со схемой и онтологией, но на практике, как правило, это не поддерживается в полной мере.

Приведем ряд примеров. В [29] предлагается разделять определения классов онтологии и их экземпляров для хранения в БД. В [31] описывается инструментальное средство для запоминания онтологий OWL в РБД с использованием конфигурационных XML-файлов. В [30] предлагается вариант OBDB, в которой расширяется классическая архитектура БД ANSI/SPARC введением семантических данных.

Подведем итоги проведенного анализа — отметим следующую специфику этих исследований, раскрыв суть предлагаемого подхода:

- отсутствие формального подхода. В качестве информационной модели семантического веба выбирается онтология и, как правило, OWL. В рассматриваемом случае для описания отображений выбрана ДЛ, что дает возможность поставить проблему отображений на более формальный уровень;

- односторонний механизм описания отображений. Как правило, независимо описываются отображения из онтологии в РМД, и наоборот. Проблема установления отображения в обе стороны в рамках единого подхода не решается. Необходимо, чтобы в предлагаемом подходе эта проблема решалась;

- отображение только структурных составляющих моделей. Существующие механизмы описывают отображение лишь структурных компонентов МД, а не конструкторов концептов и ролей, а также операций реляционной алгебры. Предлагаемый подход заключается в том, что он должен отражать и эту составную часть отображений;

- конструкции, которые не отображаются. Непонятно, каким образом отображать в классическую реляционную МД аксиомы ДЛ (равенства, включения и непересекаемости концептов). Предлагается реляционная структура, которая предоставляет такие возможности;

- гипотезы открытого мира и неуникальности имен. Дескриптивная логика строится в предположении существования гипотез открытого мира и неуникальности имен. В свою очередь, реляционная МД формулируется из предположения существования гипотез замкнутого мира и уникальности имен. Предложив вариант реляционной структуры, авторы настоящей статьи отчасти решили эти противоречия.

4. БИНАРНАЯ РЕЛЯЦИОННАЯ МОДЕЛЬ ДАННЫХ

Для решения вопроса установления взаимосвязи между ДЛ и РМД можно рассматривать классическую реляционную МД Кодда. Но в таком случае возникнет ряд проблем, способ решения которых непонятен. Это относится, например, к поддержке в РМД гипотезы открытого мира, а также конструкторов и аксиом концептов и ролей. В связи с этим определяется бинарная реляционная МД, в которой перечисленные проблемы решаются. Детально предлагаемая модель описана в [32]. Далее дано определение ее основных понятий и операций.

Бинарная реляционная модель данных (RM^2) — это совокупность бинарной реляционной структуры данных, бинарной реляционной алгебры и ограничений целостности.

Бинарная реляционная структура данных — это совокупность реляционных отношений арности не более 2. Так как в настоящей статье речь идет только о реляционных отношениях, то для краткости используем термин «отношение». Отношение — это пара (R^S, R^E) , где R^S — схема (интенционал) отношения, а R^E — экземпляр (экстенционал) отношения. Схема отношения R^S — это выражение вида $R(A_1:D_1, A_2:D_2)$, где R — имя отношения, A_1, A_2 — имена атрибутов отношения, D_1, D_2 — имена доменов отношения, определяющих области допустимых значений атрибутов.

Если указание имен доменов несущественно, то схема записывается следующим образом: $R(A_1[, A_2])$. Имена доменов и отношений являются уникальными в реляционной структуре. Имена атрибутов уникальны в схеме отношения, но могут повторяться в различных схемах.

Экземпляр отношения R^E — это подмножество либо домена, на котором определен атрибут унарного отношения $R^E \subseteq D_1$, либо декартового произведения доменов, на которых определены два атрибута бинарного отношения $R^E \subseteq D_1 \times D_2$.

Бинарная реляционная алгебра (RA^2) — это совокупность операций, определенных на бинарной реляционной структуре. Она замкнута, т.е. результат выполнения любой операции также является отношением. Имеется три вида операций:

- не увеличивающие арности результирующего отношения (являются полными аналогами операций классической реляционной алгебры (RA));
- увеличивающие арность результирующего отношения (имеют свои аналоги в RA , но в RA^2 вводятся их модификации, результирующие отношения которых не превышают арности 2);
- являющиеся новыми по отношению к RA .

Все три типа операций RA^2 рассмотрены ниже.

4.1. Операции, унаследованные из RA . К первому виду операций, которые не увеличивают арности результирующего отношения, относятся: именование, объединение, пересечение, разность, проекция, селекция, деление. Ко второму виду относятся операции декартового произведения и соединения. Как отмечалось ранее, все эти операции унаследованы из RA , а для декартова произведения и соединения приводятся соответствующие модификации, результирующие отношения которых не превышают арности 2.

Рассмотрим каждую операцию в отдельности.

Именование/переименование (ρ). Используем следующие три варианта этой операции:

- $\rho_R(E)$ — отношению, полученному в результате вычисления выражения $E RA^2$, присваивается имя R ;
- $\rho_{R(A_1[, A_2])}(E)$ — отношению, полученному в результате вычисления выражения $E RA^2$, присваивается схема $R(A_1[, A_2])$;
- $\rho_{A/B}(E)$ — в отношении, полученном в результате вычисления выражения $E RA^2$, атрибут A переименовывается в атрибут B .

Над совместимыми отношениями определяются следующие три множественные операции:

— **объединение** (\cup) $R_1 \cup R_2 = \{r \mid r \in R_1 \vee r \in R_2\}$;

— **пересечение** (\cap) $R_1 \cap R_2 = \{r \mid r \in R_1 \wedge r \in R_2\}$;

— **разность** ($-$) $R_1 - R_2 = \{r \mid r \in R_1 \wedge r \notin R_2\}$.

Проекция (π). Проекцией бинарного отношения со схемой $R(A, B)$ на атрибут A (или B), обозначаемой $\pi_A(R)$ (или $\pi_B(R)$), называется отношение, экземпляр которого содержит только A -компоненты (B -компоненты) кортежей отношения R :

$$\pi_A(R) = \{r[A] \mid r \in R\}, \quad \pi_B(R) = \{r[B] \mid r \in R\}.$$

Пусть θ — любой из следующих предикатов сравнения: $=, \neq, <, \leq, >, \geq$.

Селекция (σ). Пусть M и N — θ -сравнимые атрибуты или наборы атрибутов отношения R . Селекцией отношения R по условию $M \theta N$, обозначаемой $\sigma_{M \theta N}(R)$, называется отношение, экземпляр которого содержит те кортежи R , на которых истинно условие $M \theta N$:

$$\sigma_{M \theta N}(R) = \{r \mid r \in R \wedge r[M] \theta r[N]\}.$$

Один из атрибутов M или N может быть константой C . Например, если константой является N , то селекция принимает вид

$$\sigma_{M \theta N}(R) = \{r \mid r \in R \wedge r[M] \theta C\}.$$

В этом случае селекция также применима к унарным отношениям.

Определение 1. Пусть задано отношение $R(A, B)$. Образом кортежа $r_A(R) \in \pi_A(R)$, записанным как $Ir_A(R)$, называется такое множество кортежей $r_B(R) \in \pi_B(R)$, конкатенация которых с r_A принадлежит R :

$$Ir_A(R) = \{r_B \mid r_B \in \pi_B(R) \wedge (r_A, r_B) \in R\}.$$

Деление (\div). Пусть заданы отношения $R_1(A, B)$ и $R_2(C, D)$, причем B и C сравнимы по предикату равенства. Делением R_1 на R_2 по B и C , обозначенным $R_1[B \div C]R_2$, называется отношение, которое состоит из таких кортежей $r_A \in \pi_A(R_1)$, образы которых $Ir_A(R_1)$ содержат все кортежи проекции $\pi_C(R_2)$:

$$R_1[B \div C]R_2 = \{r_A \mid r_A \in \pi_A(R_1) \wedge Ir_A(R_1) \supseteq \pi_C(R_2)\}.$$

Деление следующим образом выражается через уже определенные операции:

$$R_1[B \div C]R_2 = \pi_A(R_1) - \pi_A((\pi_A(R_1) \times \pi_C(R_2)) - R_1).$$

(Декартово) произведение (\times). (Декартовым) произведением унарных отношений R_1 и R_2 , обозначенным $R_1 \times R_2$, называется бинарное отношение, экземпляр которого содержит все возможные конкатенации кортежей отношений R_1 и R_2 :

$$R_1 \times R_2 = \{(r_1, r_2) \mid r_1 \in R_1 \wedge r_2 \in R_2\}.$$

Произведение двух бинарных или бинарного и унарного отношений не допускаются в RM^2 .

Соединение (\bowtie). В RM данная операция увеличивает арность результата. В связи с этим в RM^2 предлагается модифицированный вариант соединения, суть которого заключается в том, что если результирующее отношение превышает арность 2, то к нему применяется проекция по одному или двум атрибутам.

Пусть $R_1(B)$ и $R_2(C)$ — унарные отношения, где B и C — θ -сравнимые атрибуты.

Соединение отношения R_1 с отношением R_2 по условию $B \theta C$, обозначенное $R_1 \bowtie_{B \theta C} R_2$, определяется следующим образом:

$$R_1 \bowtie_{B \theta C} R_2 = \{(r_1, r_2) \mid r_1 \in R_1 \wedge r_2 \in R_2 \wedge r_1[B] \theta r_2[C]\}.$$

Пусть один или оба отношения имеют арность 2. Пусть A_i, A_j — любые два атрибута из множества атрибутов исходных отношений. Пусть также R_1 содержит атрибут A_m , а R_2 содержит A_n , которые являются θ -сравнимыми, тогда соединение R_1 и R_2 по условию $A_m \theta A_n$, обозначенное $R_1 \bowtie_{\frac{[A_i, A_j]}{A_m \theta A_n}} R_2$, определяется следующим образом:

$$R_1 \bowtie_{\frac{[A_i, A_j]}{A_m \theta A_n}} R_2 = \pi_{A_i, A_j} (R_1 \bowtie_{A_m \theta A_n} R_2).$$

Также допускается многократное соединение унарных и/или бинарных отношений с последующей проекцией на два атрибута, т.е.

$$\pi_{A_i, A_j} (R_1 \bowtie_{A_m \theta A_n} R_2 \bowtie \dots \bowtie_{A_p \theta A_q} R_k).$$

4.2. Операции, вводимые только в RA^2 . Определим еще несколько операций, которых нет в RA .

Инверсное деление ($/$). Пусть заданы отношения $R_1(A, B)$ и $R_2(C, D)$, причем B и C сравнимы по предикату равенства. Инверсным делением R_1 на R_2 по B и C , обозначенным $R_1[B/C]R_2$, называется отношение, которое состоит из таких кортежей $r_A \in \pi_A(R_1)$, образы которых $Ir_A(R_1)$ содержатся среди кортежей проекции $\pi_C(R_2)$:

$$R_1[B/C]R_2 = \{r_A | r_A \in \pi_A(R_1) \wedge Ir_A(R_1) \subseteq \pi_C(R_2)\}.$$

В работе [10] доказано, что инверсное деление следующим образом выражается через уже определенные операции:

$$R_1[B/C]R_2 = \pi_A(R_1) - \pi_A(R_1 \cap (\pi_A(R_1) \times (\pi_B(R_1) - \pi_C(R_2)))).$$

Номинал ($\{\}$). Суть этой операции — построение унарного или бинарного отношения из заданной константы или пары констант. В RA такой операции нет, так как эта алгебра не рассматривает вопросы создания отношений. Далее эта операция включена в целях поддержания одноименного конструктора ДЛ. Кроме того, эта операция не выходит за рамки реализаций реляционной МД, так как все реляционные СУБД предоставляют такую возможность.

Пусть заданы константы a и b . Тогда выражения $\{a\}$ и $\{(a, b)\}$ называются номиналами и означают построение унарного (бинарного) отношения, содержащего единственный кортеж: a или (a, b) . Схема такого отношения не определена и задается операцией именованная.

Транзитивное замыкание ($+$). В RA такой операции нет. Однако эта проблема решена в реляционных СУБД, поддерживающих рекурсивный SQL. Подробнее об этом см. в [33]. Эта операция также включается в RA^2 для поддержания соответствующего конструктора ДЛ.

Пусть задано бинарное отношение со схемой $R(A, B)$. Обозначим R^+ операцию транзитивного замыкания отношения. Она определяется следующим образом:

$$R^+ = \{r | \forall r_1 \forall r_2 \in R(r_1 = (a_1, b_1) \wedge r_2 = (a_2, b_2) \wedge b_1 = a_2 \rightarrow r = (a_1, b_2))\}.$$

5. ОТОБРАЖЕНИЕ ДЛ В RM^2

В работах [10, 34–36] приводятся результаты отображения ДЛ в RM^2 . Кратко изложим их. Вначале отобразим понятия ДЛ в структуру RM^2 , а затем отобразим конструкторы и аксиомы концептов и ролей в выражения RA^2 .

5.1. Отображение ДЛ в структуру RM^2 . Для описания такого отображения вначале построим концептуальную информационную модель ДЛ, а затем преобразуем ее в RM^2 .

Одна из основных задач концептуальной информационной модели любой предметной области — определить основные понятия, описать их свойства и взаимосвязи. Одним из наиболее используемых для этих целей является язык ER, который предполагает, что концептуальная информационная структура описывается с помощью таких понятий, как сущность, атрибут и связь. В данном случае используется диалект языка ER, предложенный Баркером [37].

На рис. 2 приведена концептуальная информационная модель ДЛ. В модели имеется три базисных сущности: для представления концептов (Concept), ролей (Role) и индивидов (Individual). Каждый из них имеет единственный

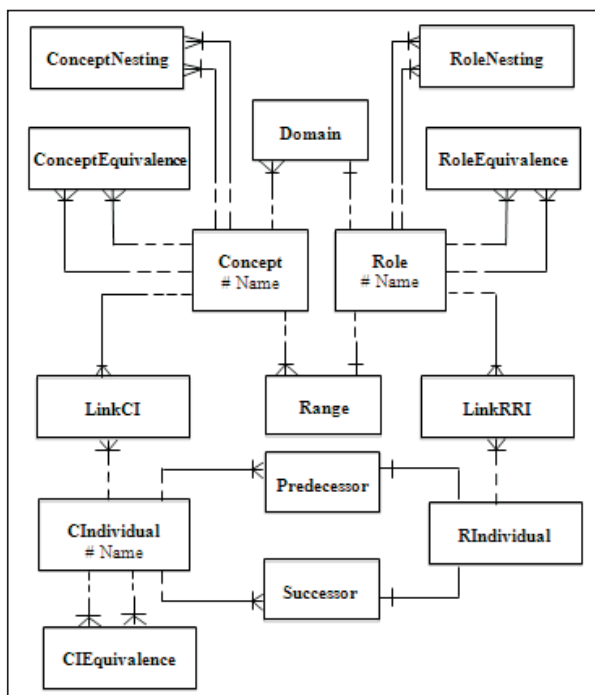


Рис. 2. Концептуальная информационная модель дескриптивной логики

атрибут Name, который уникально их идентифицирует. Все остальные сущности являются сущностями-связями, представляющими бинарные взаимосвязи между базисными сущностями. Они не имеют собственных атрибутов и уникально идентифицируются своими связями. Представим их:

- LinkCI — разрешает связь многие-ко-многим между концептами и индивидами;
- RIndividual — представляет экземпляр роли;
- LinkRRI — разрешает связь многие-ко-многим между ролями и их экземплярами;
- Domain — представляет область определения роли;
- Range — представляет область значений роли;
- Predecessor — представляет первый индивид экземпляра роли;
- Successor — представляет второй индивид экземпляра роли;
- ConceptNesting, ConceptEquivalence, RoleNesting, RoleEquivalence — представляют аксиомы включения и эквивалентности концептов и ролей;
- CIEquivalence — представляет аксиому эквивалентности индивидов.

Теперь построим схему RM^2 , которая соответствует приведенной ER-модели. При этом используем алгоритм преобразования, описанный в [38]. Кроме того, воспользуемся идеей введения атрибутов-заменителей для представления первичных ключей, которая впервые предложена основателем РМД Э. Коддом [39]. Так как построена ER-схема, соответствующая требованию 3NF, то каждая сущность представляется в виде отношения RM^2 (оставляем имя отношения таким же, как и имя соответствующей сущности):

Concept (CPK, Name)	Range (CFK, RFK)	ConceptNesting (CInFK, COutFK)
Role (RPK, Name)	LinkCI (CFK, CIfK)	ConceptEquivalence (CForFK, CIsFK)
CIndividual (CIPK, Name)	LinkRRI (RFK, RIfK)	RoleNesting (RInFK, ROutFK)
RIndividual (RIPK)	Predecessor (CIfK, RIfK)	RoleEquivalence (RForFK, RIsFK)
Domain (CFK, RFK)	Successor (CIfK, RIfK)	CIEquivalence (CIfForFK, CIIsFK)

В этой схеме атрибуты с окончанием РК являются первичными ключами, а с окончанием FK — внешними ключами.

5.2. Отображение ДЛ ALC в RM^2 . Начнем описание отображения с простейшего варианта ДЛ, названного ALC. Синтаксис ALC определяется следующим образом:

$$T | \perp | A | \neg C | C \sqcap D | C \sqcup D | \exists R.C | \forall R.C,$$

где A — атомарный концепт, R — атомарная роль, C, D — произвольные концепты.

Семантика ALC задается через понятие интерпретации. Интерпретация — это пара $I = (\Delta, \cdot^I)$, где Δ — непустое множество, называемое областью интерпретации, а \cdot^I — интерпретирующая функция, которая присваивает каждому атомарному концепту A множество $A^I \subseteq \Delta$, а каждой атомарной роли R — бинарное отношение $R^I \subseteq \Delta \times \Delta$. Остальные формулы интерпретируются следующим образом:

$$\begin{aligned} T^I &= \Delta, \perp^I = \emptyset, (\neg A)^I = \Delta \setminus A^I, \\ (C \sqcap D)^I &= C^I \cap D^I, (C \sqcup D)^I = C^I \cup D^I, \\ \exists R.C &= \{a \in \Delta \mid \exists b \in \Delta ((a, b) \in R^I \wedge b \in C^I)\}, \\ \forall R.C &= \{a \in \Delta \mid \forall b \in \Delta ((a, b) \in R^I \rightarrow b \in C^I)\}. \end{aligned}$$

В дальнейшем $C_{RM^2}^E$ обозначает экстенционал отношения RM^2 , которое ставится в соответствие интерпретации произвольного концепта C ДЛ.

Покажем, как выражаются в RM^2 атомарные концепты и роли. Для концепта C и роли R эти выражения следующие:

$$\begin{aligned} C_{RM^2}^E &= \pi_{CIndividual.Name} (\sigma_{Concept.Name='C'} (CIndividual \bowtie_{CIPK=CIFK} (LinkCI \bowtie_{CFK=CPK} Concept))), \\ R_{RM^2}^E &= \pi_{First,Second} (\rho_{CIndividual.Name/Second} (CIndividual \bowtie_{CIPK=CIFK} (Successor \bowtie_{RIFK=RIPK} (\rho_{CIndividual.Name/First} (CIndividual \bowtie_{CIPK=CIFK} (Predecessor \bowtie_{RIFK=RIPK} (\sigma_{Role.Name='R'} (RIndividual \bowtie_{RIPK=RIFK} (LinkRRI \bowtie_{RFK=RPK} Role))))))))). \end{aligned}$$

Предполагается, что концепту T соответствует отношение $CIndividual$, т.е. $T_{RM^2}^E = \pi_{Name} (CIndividual)$. А для концепта \perp в RM^2 вводится специальное отношение $Empty(Name)$, которое не содержит экземпляров. Все остальные концепты ALC представляются в RM^2 следующим образом:

$$\begin{aligned} (\neg C)_{RM^2}^E &= \pi_{Name} (CIndividual) - C_{RM^2}^E, \\ (C \sqcup D)_{RM^2}^E &= C_{RM^2}^E \cup D_{RM^2}^E, (C \sqcap D)_{RM^2}^E = C_{RM^2}^E \cap D_{RM^2}^E, \\ (\exists R.C)_{RM^2}^E &= \pi_{First} (R_{RM^2}^E \bowtie_{Second=Name} C_{RM^2}^E), \\ (\forall R.C)_{RM^2}^E &= \pi_{First} (R_{RM^2}^E) - \pi_{First} (R_{RM^2}^E \cap (\pi_{First} (R_{RM^2}^E) \times \\ &\quad \times (\pi_{Second} (R_{RM^2}^E) - \pi_{Name} (C_{RM^2}^E)))). \end{aligned}$$

5.3. Отображение расширений ДЛ ALC в RM^2 . R-последователь — это такой индивид, который является правой частью роли R. Множество R-последовательных для e обозначим $R^I(e)$, где $e \in \Delta$:

$$R^I(e) = \{d \in \Delta \mid (e, d) \in R^I\}.$$

Мощность такого множества обозначим $|R^I(e)|$.

Численные ограничения роли. Следующие конструкторы называются численными ограничениями на роли. Если R — роль, C — концепт, а $n \geq 0$ — натуральное число, то:

- $(\leq 1R)$ есть концепт — ограничение функциональности;
- $(\leq nR)$ и $(\geq nR)$ есть концепт — количественное ограничение;
- $(\leq nR.C)$ и $(\geq nR.C)$ есть концепт — качественное ограничение.

Интерпретация этих конструкторов следующая:

$$(\leq 1R)^I = \{e \in \Delta \mid |R^I(e)| \leq 1\},$$

$$(\leq nR)^I = \{e \in \Delta \mid |R^I(e)| \leq n\}, \quad (\geq nR)^I = \{e \in \Delta \mid |R^I(e)| \geq n\},$$

$$(\leq nR.C)^I = \{e \in \Delta \mid |R^I(e) \cap C^I| \leq n\}, \quad (\geq nR.C)^I = \{e \in \Delta \mid |R^I(e) \cap C^I| \geq n\}.$$

Для решения вопроса представления количественных ограничений в RM^2 воспользуемся следующей их взаимосвязью с исчислением предикатов. Пусть задана роль R и соответствующий ей предикат $P(x, y)$. Тогда

$$(\geq nR) \Leftrightarrow \exists y_1 \dots \exists y_n (P(x, y_1) \wedge \dots \wedge P(x, y_n) \wedge \bigwedge_{1 \leq i < j \leq n} y_i \neq y_j).$$

Отсюда очевидно представление такого концепта в RM^2 . Например, для $(\geq 2R)$:

$$(\geq 2R)_{RM^2}^E = \pi_{\text{First}} (\sigma_{\text{Second1} \neq \text{Second2}} (\rho_{\text{Second/Second1}} (R_{RM^2}^E) \bowtie \bowtie_{\text{First=First}} (\rho_{\text{Second/Second2}} (R_{RM^2}^E)))).$$

Таким образом, с учетом того, что концепт $(\leq 1R)$ эквивалентен $\neg(\geq 2R)$ и что $T_{RM^2}^E = \pi_{\text{Name}}(C\text{Individual})$, ограничение функциональности представляется так:

$$(\leq 1R)_{RM^2}^E = \pi_{\text{Name}}(C\text{Individual}) - (\geq 2R)_{RM^2}^E.$$

Теперь становится очевидным представление в RM^2 концепта $(\geq nR)$:

$$(\geq nR)_{RM^2}^E = \pi_{\text{First}} \left(\sigma_{\substack{1 \leq i < j \leq n \\ R_i \text{ Second}_i \neq R_j \text{ Second}_j}} \left(\overset{*}{\underset{1 \leq i \leq n}{\rho_{R_i}(\text{First}, \text{Second}_i)}} (R_{RM^2}^E) \right) \right).$$

Здесь выполняется операция переименования отношения $R_{RM^2}^E$, чтобы различать его экземпляры. Выражение $\overset{*}{\underset{1 \leq i \leq n}{\rho_{R_i}(\text{First}, \text{Second}_i)}}$ указывает на выполнение n-кратного естественного соединения отношений, а выражение $\sigma_{\substack{1 \leq i < j \leq n \\ R_i \text{ Second}_i \neq R_j \text{ Second}_j}}$ — на выполнение многократной селекции. Концепт $(\leq nR)$ эквивалентен концепту $\neg(\geq n+1R)$ и представляется в RM^2 следующим образом:

$$(\leq nR)_{RM^2}^E = \pi_{\text{Name}}(C\text{Individual}) - (\geq n+1R)_{RM^2}^E.$$

Отображения для качественного ограничения похожи на отображения количественного ограничения с одним уточнением: индивиды R-последователя должны принадлежать концепту C. Это достигается предварительным соединением

$R_{RM^2}^E$ и $C_{RM^2}^E$, в результате чего в $R_{RM^2}^E$ отбрасываются те кортежи, у которых R-последователи не принадлежат $C_{RM^2}^E$ т.е. $R_{RM^2}^E \bowtie_{\text{Second} = \text{Name}} C_{RM^2}^E$. Таким образом, для $(\geq nR.C)$ получим

$$(\geq nR.C)_{RM^2}^E = \pi_{\text{First}} \left(\sigma_{\substack{1 \leq i < j \leq n \\ R_i \text{Second}_i \neq R_j \text{Second}_j}} \left(\rho_{R_i(\text{First}, \text{Second}_i)}^* \left(R_{RM^2}^E \bowtie_{\text{Second} = \text{Name}} C_{RM^2}^E \right) \right) \right).$$

Аналогично получается для $(\leq nR.C)$. Закончив с численными ограничениями отметим, что с учетом следующих аксиом эквивалентности концептов:

$$(\equiv nR) \equiv (\geq nR) \sqcap (\leq nR), (\lt nR) \equiv (\leq nR) \sqcap \neg(= nR), (\gt nR) \equiv (\geq nR) \sqcap \neg(= nR),$$

$$(\equiv nR.C) \equiv (\geq nR.C) \sqcap (\leq nR.C), (\lt nR.C) \equiv (\leq nR.C) \sqcap \neg(= nR.C),$$

$$(\gt nR.C) \equiv (\geq nR.C) \sqcap \neg(= nR.C),$$

становится очевидным, как в RM^2 выражаются численные ограничения $=, >$ и $<$.

Номиналы. Номинал — это конструктор концепта, который строит концепт из индивида. Если d есть имя индивида, то $\{d\}$ есть концепт. Семантика номинала следующая: $\{d\}^I = \{d^I\}$. В расширенный вариант RM^2 введена операция номинала для поддержания конструктора номинала ALC. Поэтому $\{d\}_{RM^2}^E = \{d_{RM^2}^E\}$.

5.4. Отображение конструкторов ролей ДЛ в RM^2 . Существуют расширения ALC, в которые включаются конструкторы ролей.

Определение 2 (конструкторов ролей). Пусть R и S — роли, а C — концепт, тогда ролями также являются следующие выражения: R^- (обратная роль), $\neg R$ (дополнение), $R \sqcap S$ (пересечение), $R \sqcup S$ (объединение), $R \circ S$ (композиция), $\text{id}(C)$ (тождественность концепта), R^+ (транзитивное замыкание), R^* (рефлексивно-транзитивное замыкание). Их семантика следующая:

$$(R^-)^I = \{(e, d) \in \Delta \times \Delta \mid (d, e) \in R^I\}, (\neg R)^I = \Delta \times \Delta \setminus R^I,$$

$$(R \sqcap S)^I = R^I \cap S^I, (R \sqcup S)^I = R^I \cup S^I,$$

$$(R \circ S)^I = \{(e, d) \in \Delta \times \Delta \mid \exists c \in \Delta ((e, c) \in R^I \wedge (c, d) \in S^I)\},$$

$$(\text{id}(C))^I = \{(e, e) \in \Delta \times \Delta \mid e \in C^I\}.$$

Пусть $R^0 = \{(e, e) \mid e \in \Delta\}$, а R^n — $R \circ R \circ \dots \circ R$ n раз, тогда

$$(R^+)^I = \bigcup_{n \geq 1} (R^I)^n, (R^*)^I = \bigcup_{n \geq 0} (R^I)^n.$$

Если в ДЛ семантика каждого из пары индивидов в роли выражается его местом (первый индивид является R-предшественником, а второй — R-преемником), то в РМД семантика атрибутов раскрывается через их имена. В связи с этим перестановка местами индивидов в роли для получения обратной роли равносильна перестановке имен атрибутов в RM^2 . Поэтому отображение в RM^2 для концепта R^- имеет вид

$$(R^-)_{RM^2}^E = (\rho_{R(\text{Second}, \text{First})}(R_{RM^2}^E)).$$

Для остальных конструкторов ролей их представление в RM^2 очевидным образом выводится из их интерпретации:

$$\begin{aligned}
& (\neg R)_{RM^2}^E = \\
& = (\rho_{Name/First}(\pi_{Name}(CIndividual)) \times \rho_{Name/Second}(\pi_{Name}(CIndividual))) - R_{RM^2}^E, \\
& (R \sqcap S)_{RM^2}^E = R_{RM^2}^E \cap S_{RM^2}^E, (R \sqcup S)_{RM^2}^E = R_{RM^2}^E \cup S_{RM^2}^E, \\
& (R \circ S)_{RM^2}^E = \pi_{R.First, S.Second}(R_{RM^2}^E \bowtie R.Second=S.First S_{RM^2}^E), \\
& (id(C))_{RM^2}^E = (\rho_{Name/First}(\pi_{Name} C_{RM^2}^E)) \bowtie_{First=Second} (\rho_{Name/Second}(\pi_{Name} C_{RM^2}^E)), \\
& (R^+)_{RM^2}^E = (R_{RM^2}^E)^+.
\end{aligned}$$

Что касается рефлексивно-транзитивного замыкания, то учитывая следующую аксиому ДЛ $R^* \equiv id(T) \sqcup R^+$, которая показывает, как рефлексивно-транзитивное замыкание выражается через транзитивное замыкание, получаем следующее:

$$\begin{aligned}
(R^*)_{RM^2}^E & = (\rho_{Name/First}(\pi_{Name} CIndividual)) \bowtie \\
& \bowtie_{First=Second} (\rho_{Name/Second}(\pi_{Name} CIndividual)) \cup (R^+)_{RM^2}^E.
\end{aligned}$$

5.5. Отображение аксиом ДЛ в RM^2 . Отображение аксиоматики ДЛ понятно. Концептуальная ER-схема ДЛ содержит отдельную сущность для каждой аксиомы, а значит, в RM^2 -схеме имеются следующие бинарные отношения для аксиом, ссылающиеся своими внешними ключами на концепты, роли или индивиды, относительно которых формулируются аксиомы:

ConceptNesting (CInFK, COutFK)	$C \subseteq D$,
ConceptEquivalence (CForFK, CIsFK)	$C \equiv D$,
RoleNesting (RInFK, ROutFK)	$R \subseteq S$,
RoleEquivalence (RForFK, RIsFK)	$R \equiv S$,
CIEquivalence (CForFK, CIsFK)	$a = b$.

ЗАКЛЮЧЕНИЕ

Описывается бинарная реляционная МД RM^2 и предлагается механизм отображения дескриптивной логики ALC и ее расширений в эту МД. Представляет также интерес решение обратной задачи — отображения RM^2 в дескриптивную логику, которая является предметом дальнейших исследований авторов статьи.

СПИСОК ЛИТЕРАТУРЫ

1. Когаловский М.Р. Методы интеграции данных в информационных системах. Москва: Ин-т проблем рынка РАН, 2010. 74 с.
2. Чистякова И.С. Онтолого-ориентированная интеграция данных в семантическом вебе. *Проблемы программирования*. 2014. № 2–3. С. 188–196.
3. Lenzerini M. Data integration: a theoretical perspective. *Proc. of the 21st ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS 2002)*. New York: ACM Press, 2002. P. 233–246.
4. Бездушный А.А. Математическая модель системы интеграции данных на основе онтологий. *Вестник НГУ. Сер. Информационные технологии*. 2008. Т. 6, вып. 2. С. 15–40.
5. Чистякова И.С. Роль онтологий в интеграции данных в семантическом вебе. *Инженерия программного обеспечения*. 2014. № 1(17). С. 52–63.

6. Isabel F. Cruz, Huiyong Xiao. The role of ontologies in data integration. *Engineering intelligent systems for electrical engineering and communications*. 2005. N 4. P. 245–252.
7. Guarino N., Giaretta P. Ontologies and knowledge bases: towards a terminological clarification. *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*. N. Mars (Ed.). Amsterdam: IOS Press, 1995. P. 25–32..
8. The description logic handbook: theory, implementation, and applications. Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P. (Eds.) Cambridge University Press, 2003. 555 p.
9. Kevin Chen-Chuan Chang, Bin He, Chengkai Li, Mitesh Patel, Zhen Zhang. Structured databases on the web: observations and implications. *ACM SIGMOD Record*. 2004. Vol. 33, N 3. P. 61–70.
10. Резниченко В.А., Чистякова И.С. Отображение дескриптивной логики ALC в бинарную реляционную структуру данных. *Проблемы програмування*. 2015. № 4. С. 13–30.
11. Vysniauskas E., Nemuraite L.. Transforming ontology representation from OWL to relational database. *Information Technology and Control*. 2006. Vol 35, N 3A. P. 333–343.
12. Martinez-Cruz C., Blanco I.J., Vila M.A. Ontologies versus relational databases: are they so different? A comparison. *Artificial Intelligence Review*. 2012. Vol. 38, N 4. P. 271–290.
13. Brockmans S., Haase P., Hitzler P., Studer R. A metamodel and UML profile for rule-extended OWL DL ontologies. *The Semantic Web: Research and Applications*. Sure Y., Domingue J. (eds). ESWC 2006. *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg. 2006. Vol 4011. P. 303–316.
14. Colomb R.M., Gerber A., Lawely M. Issues in mapping metamodels in the ontology development metamodel using QVT. *EDOC workshop on Model-Driven Semantic Web (MDSW)*. Monterey, USA. September 2004. 2004. P. 20–24.
15. Upadhyaya S.R., Kumar P.S. ERONTO: A tool for extracting ontologies from extended ER diagrams. *Proc. of the 2005 ACM Symposium on Applied Computing (SAC)*. 2005. P. 666–670.
16. Astrova I., Korda N., Kalja A. Storing OW ontologies in SQL relational databases. *World academy of science, engineering and technology. International Journal of Computer, Electrical, Automation, Control and Information Engineering*. 2007. Vol. 1, N 5. P. 1261–1266.
17. El-Ghalayini H., Odeh M., McClatchey R. Engineering conceptual data models from domain ontologies: a critical evaluation. *International Journal of Information Technology and Web Engineering (IJITWE)*. 2006. Vol. 2, N 1. P. 57–70 .
18. Juric D., Skocir Z. Building OWL ontologies by analyzing relational database schema concepts and WordNet semantic relations. Juric D., Skocir Z. (eds). *Proc. of the 9th International Conference on Telecommunications, ConTEL*. Zagreb, Croatia. 2007. P. 235–242.
19. Stojanovic L., Stojanovic N., Volz R. Migrating data-intensive web sites into the semantic web. *SAC'02: Proc. of the 2002 ACM Symposium on Applied Computing*. ACM, New York. 2002. P. 1100–1107.
20. Champin P.A., Houben G., Thiran P. Cross: an OWL wrapper for reasoning on relational databases. *Proc. of the 26th International Conference on Conceptual Modeling*. Springer, Auckland, New Zealand. 2007. P. 502–517.
21. Doan A., Madhavan J., Domingos P., Havley A. Learning to map between ontologies on the semantic web. *Proc. of the 11th International Conference on World Wide Web*. WWW. 2002. P. 662–673.
22. Sonia D., Khan S. R2O transformation system: relation to ontology transformation for scalable data integration. *IDEAS '08: Proc. of the 2008 International Symposium on Database Engineering and Applications*. ACM. New York, 2008. P. 291–295.
23. Astrova I., Kalja A. Mapping of SQL relational schemata to OWL ontologies. *Proc. of the 6th WSEAS International Conference on Applied Informatics and Communications*. 2006. P. 375–380.
24. Astrova I. Towards the semantic web — an approach to reverse engineering of relational databases to ontologies. *Advances in Databases and Information Systems: 9th East-European Conference*. 2005. P. 111–122.
25. Shuefeng Zhou, Haiyun Ling, Mei Han, Huaiwei Zhang. Ontology generator from relational database based on jena. *Computer and Information Science*. 2010. Vol. 3, N 2. P. 263–267.
26. Shuefeng Zhou, Guangwu Meng, Haiyun Ling. Ontology acquisition from relational database. *Computer and Information Science*, 2010. Vol. 3, N 1. P. 185–187.
27. Lei Zhang, Jing Li. Automatic generation of ontology based on database. *Journal of Computational Information Systems*. 2011. Vol. 7, N 4. P. 1148–1154.
28. Kashyap V. Design and creation of ontologies for environmental information retrieval. *Proc. of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management*. Voyager Inn, Banff, Alberta, Canada, 1999. p.

29. Pierra J.S., Ait-Ameur Y. Domain ontologies: a database-oriented analysis. *Proc. of the Web Information Systems and Technologies*. 2006. P. 238–254.
30. Fankam C., Jean S., Bellatreche L., Ait-Ameur Y. Extending the ANSI/SPARC architecture database with explicit data semantics: an ontology-based approach. *ECSA '08: Proc. of the 2nd European Conference on Software Architecture*. Berlin: Springer, 2008. P. 318–321.
31. Maria del Mar Roldan-Garcia, Jose F. Aldana-Montes. A tool for storing owl using database technology. *Proc. of the OWLED '05 Workshop on OWL: Experiences and Directions*. 2005, Vol. 188. P. 1–10.
32. Резниченко В.А., Чистякова И.С. Бинарная реляционная модель данных. *Проблемы програмування*. 2017. № 2. С. 96–105.
33. Резниченко В.А. Рекурсивный SQL. *Інженерія програмного забезпечення*. 2010. № 4. С. 63–81.
34. Резниченко В.А., Чистякова И.С. Интеграция семейства расширенных дескриптивных логик с реляционной моделью данных. *Проблемы програмування*. 2016. № 2–3. С. 38–47.
35. Чистякова И.С. Интеграция логик с операциями над ролями с реляционной моделью данных. *Проблемы програмування*. 2016. № 4. С. 58–65.
36. Чистякова И.С. Интеграция аксиоматики дескриптивных логик с реляционной моделью данных. *Проблемы програмування*. 2017. № 1. С. 51–58.
37. Barker R. Case* method: entity relationship modelling. Addison-Wesley, 1990. 240 p.
38. Пасічник В.В., Резніченко В.А. Організація баз даних та знань. Київ: Видавнича група BHV, 2006. 384 с.
39. Codd E.F. Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems (TODS)*. 1979. Vol. 4, Issue 4. P. 397–434.

Надійшла до редакції 30.06.2017

П.І. Андон, В.А. Резніченко, І.С. Чистякова
ВІДОБРАЖЕННЯ ДЕСКРИПТИВНОЇ ЛОГІКИ У РЕЛЯЦІЙНУ МОДЕЛЬ ДАНИХ

Анотація. Обговорено загальні задачі проблеми інтеграції даних, а також онтолого-орієнтовану інтеграцію даних у семантичному вебі. Показано важливість встановлення відображень між онтологіями і реляційними базами даних, наведено короткий аналітичний огляд досліджень з цієї тематики. Визначено бінарну реляційну модель даних і описано відображення дескриптивної логіки ALC та її розширень у цю модель.

Ключові слова: інтеграція даних, онтологія, семантичний веб, дескриптивна логіка, бінарна реляційна модель даних.

P.I. Andon, V.A. Reznichenko, I.S. Chistyakova
MAPPING OF DESCRIPTION LOGIC TO THE RELATIONAL DATA MODEL

Abstracts. The general topics of the data integration problem are discussed, including ontology-oriented data integration in the semantic web. The importance of establishing mappings between ontologies and relational databases is indicated, and a brief analytical review of researches on this topic is given. The binary relational data model is defined and the mapping of the ALC description logic and its extensions to the binary relational data model is described.

Keywords: data integration, ontology, semantic web, description logic, binary relational data model.

Андон Филипп Илларионович,
 академик НАН Украины, доктор физ-мат. наук, директор Института программных систем НАН Украины, Киев, e-mail: iss@isofts.kiev.ua.

Резниченко Валерий Анатольевич,
 канд. физ-мат.-наук, старший научный сотрудник, заместитель заведующего отделом Института программных систем НАН Украины, Киев, e-mail: reznich@isofts.kiev.ua.

Чистякова Инна Сергеевна,
 младший научный сотрудник Института программных систем НАН Украины, Киев, e-mail: inna_islyamova@ukr.net.