

ОТОБРАЖЕНИЕ РЕЛЯЦИОННОЙ АЛГЕБРЫ В ДЕСКРИПТИВНУЮ ЛОГИКУ

И.С. Чистякова

Работа посвящена проблеме интеграции данных, а именно отображению реляционной модели данных в дескриптивную логику. С помощью бинарной реляционной модели данных, описание которой было приведено в ранее опубликованных исследованиях, рассматривается механизм создания отображений операций реляционной алгебры в дескриптивную логику. При создании механизмов отображений в работе используется дескриптивная логика ALC и некоторые её расширения. В работе используются ранее опубликованные результаты, а именно: определение и структура бинарной реляционной модели данных, отображения дескриптивной логики ALC, её аксиоматики и расширений в бинарную реляционную модель данных.

Ключевые слова: бинарная реляционная модель данных, реляционная алгебра, дескриптивная логика, отображение данных, ALC, RDM, RM^2 , RA.

Робота присвячена проблемі інтеграції даних, а саме відображенню реляційної моделі даних в дескриптивну логіку. За допомогою бінарної реляційної моделі даних, опис якої було наведено у раніше опублікованих дослідженнях, розглядається механізм створення відображень операцій реляційної алгебри у дескриптивну логіку. Для створення механізмів відображення в роботі використовується дескриптивна логіка ALC та деякі її розширення. У роботі використовуються раніше опубліковані результати, а саме: визначення та структура бінарної реляційної моделі даних, відображення дескриптивної логіки ALC, її аксиоматики та розширень у бінарну реляційну модель даних.

Ключові слова: бінарна реляційна модель даних, реляційна алгебра, дескриптивна логіка, відображення даних, ALC, RDM, RM^2 , RA.

This paper is dedicated to the data integration problem. In article we discuss the problem of mapping the relational data model to the description logic. With the help of the previously created binary relational data model we created the mappings of the relational algebra to the description logic. We used the ALC description logic and its basic extensions (on which SHIQ, SHION and SHOIQ logic is based) to create mappings. We used previously published results in this work which are the binary relational data model definition, the binary relational data structure definition, the mapping of the ALC logic to the relational data model, the mapping of the basic extensions of the ALC logic to the relational data model and mapping the ALC axiomatic to the relational data model.

Key words: binary relational data model, relational algebra, description logic, data mapping, ALC, RDM, RM^2 , RA.

Введение

В современном научном пространстве, начиная с середины 70-х годов, существует глобальная проблема интеграции данных. Именно во второй половине прошлого века начинает формироваться представление о многоуровневой архитектуре баз данных (БД), о модели данных как инструменте информационного моделирования реальности и об отображениях между существующими моделями. Ввиду этого, акцент делался на поддержке глобальной схемы (впоследствии получившей название концептуальной) и локальных схем, в основе которых лежат различные (или одна и та же) модели данных (МД), функционирующих в разных узлах системы, которая находится под управлением одной и той же системы управления базами данных (СУБД). В дальнейшем проблему обобщили, и она стала включать в себя также создание хранилищ данных, репозиториях информационных ресурсов, веб-приложений. Сегодня ключевым аспектом интеграции данных является интеграция неоднородных данных различных информационных систем.

Как отмечалось в работе [1], объем информации, существующей в интернет-пространстве, с каждым днем возрастает в геометрической прогрессии, а её релевантность – в арифметической. Это порождает ряд проблем, связанных с использованием и хранением данных. В частности выделяют ряд причин [1], порождающих их гетерогенность, ввиду которых интеграция и совместное использование данных из множества таких источников является сложной и неизменно актуальной задачей на протяжении последних десятилетий.

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

Работа [2] посвящена подробному изучению комплексной проблемы интеграции данных, согласно которой можно выделить три основные составляющие данной проблемы:

- схемы интеграции данных;
- отображения между моделями данных;

- способы манипулирования данными.

В рамках решения комплексной проблемы интеграции данных ряд исследований [1, 3–7] был посвящен созданию отображений между моделями, а именно между дескриптивной логикой (DL), которую можно рассматривать как модель данных [3, 7] и реляционной моделью данных. В данных публикациях была представлена бинарная реляционная модель данных (RM²) – специальный вариант реляционной модели данных – промежуточная модель, с помощью которой все компоненты дескриптивной логики ALC (лежащей в основе множества современных DL), а также её основные расширения и аксиоматику, можно отобразить в реляционную модель данных.

Однако, эти исследования остаются неполными до тех пор, пока открыт вопрос возможности обратного отображения РМД в DL. Данная статья восполняет этот пробел и рассматривает вопросы отображения реляционной структуры данных и операций реляционной алгебры в дескриптивную логику.

Раздел 1 посвящен краткому описанию RM² и её составляющих. В разделе 2 приводится описание отображений операций реляционной алгебры (РА) в DL. Завершают текущую публикацию основные выводы и дальнейшие планы исследований.

Краткое описание бинарной реляционной модели данных

Для решения вопроса установления взаимосвязей между DL и РМД планировалось рассматривать классическую РМД Кодда [8]. Однако, это порождает ряд проблем, решить которые затруднительно. Например, неясно каким образом в РМД следует поддерживать гипотезу открытого мира, а также конструкторы концептов и ролей любой DL. В связи с этим в ряде работ была определена бинарная реляционная модель данных (БРМД), в которой решаются перечисленные проблемы. Её подробное описание можно найти в работах [1, 7].

Акцентируем внимание лишь на основных определениях.

Бинарная реляционная модель данных (RM²) – 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 идёт только о реляционных отношениях, то для краткости будем использовать 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 (на одном домене может быть определено множество атрибутов).

Если указание имен доменов не существенно, то схема записывается следующим образом $R(A_1 [, A_2])$.

Имена доменов и имена отношений являются уникальными в реляционной структуре. *Имена атрибутов* являются уникальными 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 (RA²) – это совокупность операций, определенных на бинарной реляционной структуре.

Она является замкнутой в том смысле, что результат выполнения любой операции также является отношением. Можно выделить три вида операций:

1) *операции, которые не увеличивают арность результирующего отношения.* Они являются полными аналогами операций классической реляционной алгебры (RA);

2) *операции, увеличивающие арность результирующего отношения.* Они имеют свои аналоги в RA. В связи с этим, в бинарной реляционной алгебре вводятся их модификации, результирующие отношения которых не превышают арности 2;

3) *операции, являющиеся новыми по отношению к RA.*

Все три типа операций RA² подробнее будут рассмотрены далее.

Операции, унаследованные из RA

Операции реляционной алгебры делятся на две большие группы: **основные** и **дополнительные**.

Основные:

1) не увеличивают арность результирующего отношения

- теоретико-множественные:

- объединение
- пересечение
- разность

- проекция

- селекция

- деление

2) увеличивают арность результирующего отношения:

- соединение

- декартово произведение.

Дополнительные:

- переименование

- присвоение

- обобщенная проекция

- внешнее соединение и т. д.

В данной статье мы будем рассматривать отображения только тех операций, которые дублируются в бинарной реляционной алгебре (RA^2). К ним относятся все операции из группы основных и дополнительная операция переименования.

Рассмотрим каждую в отдельности.

Определение 1. Два отношения со схемами $R_1(A_1 [, A_2])$ и $R_2(B_1 [, B_2])$ являются совместимыми (по объединению), если имеют одинаковую арность, и пары атрибутов (A_i, B_i) , $i = 1, 2$ определены на одном и том же домене.

Над совместимыми отношениями определяются следующие три множественные операции.

1. Объединение (\cup)

$$R_1 \cup R_2 = \{r | r \in R_1 \vee r \in R_2\}.$$

2. Пересечение (\cap)

$$R_1 \cap R_2 = \{r | r \in R_1 \wedge r \in R_2\}.$$

3. Разность ($-$)

$$R_1 - R_2 = \{r | r \in R_1 \wedge r \notin R_2\}.$$

4. 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 :

$$\pi_A(R) = \{r[A] | r \in R\},$$

$$\pi_B(R) = \{r[B] | r \in R\}.$$

Здесь запись $r[A]$ ($r[B]$) обозначает выделение из кортежа r компоненты, принадлежащей значению атрибута A (B). Схема результирующего отношения наследует имя атрибута A (B), а имя отношения определяется операцией именованья.

Определение 2. Пусть θ – любой из следующих предикатов сравнения: $=, \neq, <, \leq, >, \geq$. Атрибуты A и B одного и того же или различных отношений называются θ -сравнимыми, если для любых пар значений $a \in A$ и $b \in B$ определено выражение $a \theta b$. Наборы атрибутов $M = (A_1, A_2)$ и $N = (B_1, B_2)$ называются θ -сравнимыми, если пары атрибутов (A_i, B_i) являются θ -сравнимыми. В этом случае $M \theta N$ подразумевает следующее:

$$M \theta N = A_1 \theta B_1 \wedge A_2 \theta B_2.$$

5. 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 $M \theta N$:

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

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

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

В этом случае селекция также применима к унарным отношениям.

Схема результирующего отношения наследует имена атрибутов отношения R , а имя отношения определяется операцией именованя.

Определение 3. Пусть задано отношение $R(A,B)$. *Образом* кортежа $r_A \in \pi_A(R)$, записываемым как $I_{r_A}(R)$, называется такое множество кортежей $r_B \in \pi_B(R)$, соединение которых с r_A принадлежит R :

$$I_{r_A}(R) = \{r_B \mid r_B \in \pi_B(R) \square (r_A, r_B) \in R\}.$$

6. 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 , обозначаемым $R_1[B\div C]R_2$, называется отношение, которое состоит из таких кортежей $r_A \in \pi_A(R_1)$, образы которых $I_{r_A}(R_1)$ содержат все кортежи проекции $\pi_C(R_2)$:

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

Отношение R_2 может быть унарным. Если, например, $R_2(C)$, то:

$$R_1[B\div C]R_2 = \{r_A \mid r_A \in \pi_A(R_1) \square I_{r_A}(R_1) \square R_2\}.$$

Схема результирующего отношения наследует имя атрибута A из R_1 , а имя отношения определяется операцией именованя.

Деление следующим образом выражается через произведение, разность и проекцию:

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

7. 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 оказывается арности выше двух, 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.

Определение 4. *Соединение* отношения R_1 с отношением R_2 по условию $B \theta C$, обозначаемое $R_1 \square_{B\theta C} R_2$, определяется следующим образом:

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

Определение 5. Пусть один или оба отношения имеют арность 2, тогда соединение определяется следующим образом. Пусть A_i, A_j – любые два атрибута из множества атрибутов исходных отношений. Пусть также R_1 содержит атрибут A_m , а R_2 – A_n , которые являются θ -сравнимыми, тогда соединение R_1 и R_2 по условию $A_m \theta A_n$, обозначаемое через $R_1 \bowtie_{A_m \theta A_n}^{[A_i, A_j]} R_2$, определяется следующим образом:

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

Также допускается многократное соединение унарных и/или бинарных отношений, с последующей проекцией на два атрибута, то есть:

$$\pi_{A_1, A_j}(R_1 \bowtie_{A_m \theta A_n} R_2 \bowtie \dots \bowtie_{A_p \theta A_q} R_k).$$

Схема результирующего отношения наследует имена атрибутов отношений R_1 и R_2 , а имя отношения определяется операцией именованя.

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

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

Схема результирующего отношения наследует имена атрибутов отношений R_1 и R_2 , а имя отношения определяется операцией именованя. Произведение двух бинарных или бинарного и унарного отношений не допускаются в RM^2 .

9. Именовани/переименование (ρ). Мы будем использовать следующие три варианта этой операции:

- $\rho_R(E)$ – отношению, полученному в результате вычисления выражения $E \in RA^2$, присваивается имя R .
- $\rho_R(A1[,A2])(E)$ – отношению, полученному в результате вычисления выражения $E \in RA^2$, присваивается схема $R(A1 [, A2])$.
- $\rho_{A/B}(E)$ – в отношении, полученном в результате вычисления выражения $E \in RA^2$, атрибут A переименовывается в атрибут B .

Операции, вводимые только в RA^2

Определим еще несколько операций, которых нет в RA .

10. 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 , обозначаемым $R_1[B/C]R_2$, называется отношение, которое состоит из таких кортежей $r_A \in \pi_A(R_1)$, образы которых $I_{r_A}(R_1)$ содержатся среди кортежей проекции $\pi_C(R_2)$:

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

Отличие инверсного деления от обычного заключается в том, что в первом случае образы содержатся во втором операнде, а во втором случае – содержат второй операнд. В работах [9, 10] доказано, что инверсное деление следующим образом выражается через уже определенные операции:

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

11. Номинал ($\{\}$). Суть этой операции – построение унарного или бинарного отношения из заданной константы или пары констант. В RA такой операции нет, так как эта алгебра вообще не рассматривает вопросы создания отношений. Мы же включили эту операцию для поддержания одноименного конструктора в DL . Кроме того, эта операция не выходит за рамки реализаций реляционной модели, 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 именованя.

В работе [4] рассмотрена операция транзитивного замыкания ($+$), расширяющая стандартный синтаксис логики DL ALC. В реляционной алгебре такой операции нет. Однако, эта проблема решена в реляционных СУБД, поддерживающих рекурсивный SQL, где есть возможность выразить транзитивное замыкание. Подробнее об этом можно познакомиться в работе [6]. В связи с этим мы вводим в RA^2 аналогичную операцию *транзитивного замыкания* R^+ .

12. Транзитивное замыкание (+).

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

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

Описание отображений бинарной реляционной модели данных в дескриптивную логику

При создании бинарной реляционной модели данных мы работали с логикой ALC и её расширениями. Обоснование такого решения приведено в [1]. Очевидно, что для создания отображений РМД в DL этой логики будет недостаточно, т.к. необходимо работать с теми же расширениями, и показать их отображение в обратную сторону. Поэтому, мы остановились на дескриптивной логике SHOIQ, которая включает в свой состав все основные расширения ALC.

Дескриптивная логика ALC и её расширение

Перед созданием отображений, следует напомнить основные составляющие дескриптивной логики ALC. Её синтаксис определяется следующим образом:

$$\Box \mid \Box \mid A \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \Box R.C \mid \Box R.C,$$

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, \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:

$$\Box^I = \Delta, \quad \Box^I = \emptyset, \quad (\neg A)^I = \Delta \setminus A^I,$$

$$(C \sqcap D)^I = C^I \cap D^I,$$

$$(C \sqcup D)^I = C^I \cup D^I,$$

$$\Box R.C = \{a \in \Delta \mid \exists b \in \Delta ((a, b) \in R^I \wedge b \in C^I)\},$$

$$\forall R.C = \{a \in \Delta \mid \forall b \in \Delta ((a, b) \in R^I \rightarrow b \in C^I)\}.$$

Существуют различные расширения ALC. Приведем то из них, которое будет использоваться при описании отображения RM^2 в DL

Номинал. Номинал – это конструктор концепта, который строит концепт из индивида. Если d есть имя индивида то $\{d\}$ есть концепт. Семантика номинала следующая: $\{d\}^I = \{d^I\}$.

Универсальная роль. Содержательно универсальная роль $\Box \Box$ – это объединение всех возможных ролей, а формально она интерпретируется следующим образом: $\Box \Box^I = \Delta \times \Delta$

Конструкторы ролей. Пусть R и S являются ролями, 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), $id(C)$ (идентичность концепта), R^+ (транзитивное замыкание), R^* (рефлексивно-транзитивное замыкание).

Их семантика следующая:

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

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

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

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

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

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

Правила эквивалентного преобразования

Приведем ряд правил эквивалентного преобразования операций реляционной алгебры, которые мы будем использовать в дальнейшем. Два выражения реляционной алгебры E_1, E_2 эквивалентны, если относительно любого набора правильно построенной реляционной базы данных они дают отношения с одинаковыми экземплярами. Эквивалентность будем записывать в виде правил эквивалентности вида $E_1 \equiv E_2$.

Правила эквивалентного преобразования выглядят следующим образом.

1. Коммутативность произведения

$$E_1 \times E_2 \equiv E_2 \times E_1.$$

2. Ассоциативность произведения

$$(E_1 \times E_2) \times E_3 \equiv E_1 \times (E_2 \times E_3).$$

3. Представление произведения и селекции в виде соединения

$$\sigma_p(E_1 \times E_2) \equiv E_1 \sqcap_p E_2.$$

4. Коммутативность селекции.

Пусть p_1, p_2 – предикаты, тогда

$$\sigma_{p_1}(\sigma_{p_2}(E)) \equiv \sigma_{p_2}(\sigma_{p_1}(E)).$$

5. Коммутативность соединения

$$E_1 \sqcap_p E_2 \equiv E_2 \sqcap_p E_1.$$

6. Ассоциативность соединения

$$(E_1 \sqcap_p E_2) \sqcap_{q \& r} E_3 \equiv (E_1 \sqcap_{p \& q} (E_2 \sqcap_r E_3)),$$

где r содержит атрибуты только из E_2 и E_3 .

7. Поглощение проекций.

При наличии последовательности проекций необходима только самая внешняя, остальные могут быть опущены

$$\pi_A(\pi_B(\dots \pi_C(E) \dots)) \equiv \pi_A(E).$$

8. Дистрибутивность проекции и произведения:

а) Если набор атрибутов A принадлежит E_1 , то

$$\pi_A(E_1 \times E_2) \equiv \pi_A(E_1),$$

б) Если наборы атрибутов A и B принадлежат E_1 и E_2 соответственно, то

$$\pi_{A,B}(E_1 \times E_2) \equiv \pi_A(E_1) \times \pi_B(E_2).$$

9. Дистрибутивность проекции и соединения

$$\pi_{A,B}(E_1 \sqcap_p E_2) \equiv \pi_A(E_1) \sqcap_p \pi_B(E_2),$$

если предикат p содержит только атрибуты из $(A \square B)$, и A (B) содержат только атрибуты из E_1 (E_2).

Отображение операций бинарной реляционной алгебры в DL ALC и ее расширение

Перейдем к описанию отображений операций бинарной реляционной модели данных в дескриптивную логику. Отметим, что в РА операции селекции и соединения используют широкий набор предикатов, так называемые θ -предикаты, $\theta = \{=, \neq, <, \leq, >, \geq\}$. В нашем случае мы рассмотрим только один предикат – равенства ($=$). Такую реляционную алгебру мы назовем бинарной реляционной алгеброй с равенством и будем обозначать так: $РА^{2=}$.

Пусть заданы следующие унарные и бинарные отношения:

$$P_1(F), P_2(G), Q_1(K,L), Q_2(M,N).$$

Согласно определению выше, реляционное отношение состоит из интенционала и экстенционала. Отображение **унарного отношения** в дескриптивную логику происходит исходя из следующих правил:

- имя отношения отображается в имя *концепта*;
- экстенционал отношения отображается в *интерпретацию концепта*.

Отображение **бинарного отношения** в дескриптивную логику происходит исходя из следующих правил:

- имя отношения отображается в имя *роли*;
- экстенционал отношения отображается в *интерпретацию роли*.

Пусть также имеются атомарные концепты C, D и атомарные роли R и S дескриптивной логики.

Отображение между атомарными отношениями, концептами и ролями задается следующим образом:

$$P_1 \leftrightarrow C, \text{ если } P_1^E = C^I,$$

$$P_2 \leftrightarrow D, \text{ если } P_2^E = D^I,$$

$$Q_1 \leftrightarrow R, \text{ если } Q_1^E = R^I,$$

$$Q_2 \leftrightarrow S, \text{ если } Q_2^E = S^I.$$

Здесь верхний индекс E указывает на экстенционал отношения, а верхний индекс I – на интерпретацию концепта или роли.

Теперь рассмотрим отображения операций $РА^{2=}$.

1. Объединение

Отображение объединения двух концептов выглядит следующим образом:

$$P_1 \square P_2 \leftrightarrow C \square D.$$

Отображение объединения двух ролей выглядит следующим образом:

$$Q_1 \square Q_2 \leftrightarrow R \square S.$$

2. Пересечение

Отображение пересечения двух концептов выглядит следующим образом:

$$P_1 \square P_2 \leftrightarrow C \square D.$$

Отображение пересечения двух ролей выглядит следующим образом:

$$Q_1 \square Q_2 \leftrightarrow R \square S.$$

3. Разность

Поскольку в РА не используется теоретико-множественная операция дополнения, а в дескриптивной логике она присутствует в виде дополнения концепта, то отображение интенционала операции разности выглядит следующим образом:

$$P_1 - P_2 \leftrightarrow C \square (\neg D).$$

$$Q_1 - Q_2 \leftrightarrow R \square (\neg S).$$

4. Проекция

Операция проекции существует только для *бинарных* отношений. Исходя из определения проекции, мы получаем в результате совокупность таких a , для которых **существует** b такой, что пара (a, b) состоит в отношении R . А это соответствует определению концепта существования на всех области определения. Таким образом, отображение операции проекции выглядит следующим образом:

$$\pi_K Q_1(K, L) \leftrightarrow \Box R. \Box,$$

$$\pi_L Q_1(K, L) \leftrightarrow \Box R^- . \Box.$$

5. Селекция

Операция селекция предполагает отбор строк, удовлетворяющих некоторому условию, которое в свою очередь задается на атрибутах. В нашем исследовании отображение селекции мы будем рассматривать только по условию равенства, т.к. в дескриптивных логиках (ALC и некоторых её расширениях) существует только условие равенства концептов, ролей и индивидов.

Рассмотрим варианты такой селекции для унарных и бинарных отношений. Если один из атрибутов условия селекции является константой, то операция селекции возможна как для *унарных*, так и для *бинарных* отношений. В этом случае в результате селекции **бинарного** отношения получается такое a (совокупность таких a), которое принадлежит множеству, состоящему из константы (перечня констант) условия.

Следует напомнить, что **унарное** отношение отображается в концепт в дескриптивной логике. В результате отображения операции селекции по условию равенства мы также получаем концепт. Поскольку, в дескриптивной логике не существует операции создания концепта из концепта, но существует операция создания роли из концепта – $id(C)$, которую мы называли **идентичность концепта** – то отображение операции селекции для **унарного** отношения будет происходить следующим образом. Из концепта, который является интерпретацией унарного отношения, создаётся роль с помощью операции $id(C)$, а затем выполняется селекция по заданному условию равенства.

Отображение операции селекции, *бинарного* отношения по условию равенства атрибутов выглядит следующим образом:

$$\sigma_{K=L} Q_1(K, L) \leftrightarrow id(\Box R. \Box) \Box R.$$

Отображение операции селекции, *бинарного* отношения по условию равенства атрибута константе выглядит следующим образом:

$$\sigma_{L=1} Q_1(K, L) \leftrightarrow id(\Box R. \{1\}) \circ \Box \Box \circ id(\{1\}),$$

$$\sigma_{K=k} Q_1(K, L) \leftrightarrow id(\Box R^- . \{k\}) \circ \Box \Box \circ id(\{k\}).$$

Отображение операции селекции, *унарного* отношения по условию равенства атрибута константе выглядит следующим образом:

$$\sigma_{F=f} P_1(F) \leftrightarrow \Box (id(R)). \{f\}.$$

6. Деление

Операция обычного деления в классической RA выразима через совокупность других операций RA. Отображение операции обычного деления выглядит следующим образом:

$$Q_1(K, L)[L \div F] P_1(F) = \pi_K Q_1 - \pi_K((\pi_K Q_1 \times P_1) - Q_1) \leftrightarrow \Box R. \Box - \Box((id(\Box R. \Box) \circ \Box \Box \circ id(C)) - R). \Box.$$

7. Соединение

Рассмотрим следующие варианты операции соединения.

7.1. *Соединение двух унарных отношений* в результате дают бинарное отношение. Отображение такого варианта операции соединения выглядит следующим образом:

$$P_1(F) \Box_{F=G} P_2(G) \leftrightarrow id(C \Box D).$$

7.2. *Соединение унарного и бинарного отношения* дают в результате трехарное отношение. Поскольку в RA^2 допустимы только бинарные отношения, то нам необходимо выполнить проекцию на один или два атрибута, чтобы получить допустимое результирующее отношение. Рассмотрим следующие варианты соединения:

$$a) \pi_F(P_1(F) \Box_{F=K} Q_1(K, L)) = \pi_F(P_1(F) \Box_{F=K} \pi_K Q_1(K, L)) \leftrightarrow C \Box \Box R. \Box;$$

$$b) \pi_K(P_1(F) \square_{F=K} Q_1(K,L)) = \pi_K(P_1(F) \square_{F=K} \pi_K Q_1(K,L)) \leftrightarrow C \square \square R. \square;$$

$$c) \pi_L(P_1(F) \square_{F=K} Q_1(K,L)) \leftrightarrow \square R^-. (C \square \square R. \square);$$

$$d) \pi_{F,K}(P_1(F) \square_{F=K} Q_1(K,L)) \leftrightarrow \text{id}(C \square \square R. \square);$$

$$e) \pi_{K,L}(P_1(F) \square_{F=K} Q_1(K,L)) \leftrightarrow R - (\text{id}(\square R. \square - C) \circ \square \square \circ \text{id}(\square R^-. \square));$$

f) $\pi_{F,L}(P_1(F) \square_{F=K} Q_1(K,L))$ – равносильно формуле 7.2.e).

7.3. Соединение двух бинарных отношений даёт четырехарное отношение, которое необходимо, как и в предыдущем варианте, спроецировать для получения допустимого в RA^2 результирующего отношения.

$$Q_1(K,L) \square_{L=M} Q_2(M,N).$$

Как и выше, характер отображения зависит от варианта проекции результата соединения.

a) Если в результирующей проекции нет атрибута K, то заменяем $Q_1(K,L)$ на $\pi_L Q_1(K,L)$ и все сводится к варианту 7.2.

b) Если в результирующей проекции есть атрибут K, то получаем следующие варианты:

- $\pi_K(Q_1(K,L) \square_{L=M} Q_2(M,N))$ – заменим $Q_2(M,N)$ на $\pi_M Q_2(M,N)$, и тем самым, свели к варианту 7.2.c);
- $\pi_{K,L}(Q_1(K,L) \square_{L=M} Q_2(M,N))$ – заменим $Q_2(M,N)$ на $\pi_M(M,N)$, и тем самым, свели к варианту 7.2.e);
- $\pi_{K,M}(Q_1(K,L) \square_{L=M} Q_2(M,N))$ – сводится к варианту 7.2.b);
- $\pi_{K,N}(Q_1(K,L) \square_{L=M} Q_2(M,N)) \leftrightarrow R \circ S.$

8. Декартово произведение

Рассмотрим следующие варианты операции декартового произведения.

8.1. *Произведение двух унарных отношений.* Опираясь на [12], отображения для декартового произведения двух унарных отношений выглядят следующим образом:

$$P_1 \times P_2 \leftrightarrow \text{id}(C) \circ \square \square \circ \text{id}(D).$$

8.2. *Произведение унарного и бинарного отношений.* Произведение унарного и бинарного отношений даёт трехарное отношение, которое необходимо спроецировать на один или два атрибута, чтобы получить допустимое результирующее отношение. Рассмотрим возможные варианты проекции.

$$a) \pi_F(P_1(F) \times Q_1(K,L)) = \pi_F P_1(F) \leftrightarrow C;$$

$$b) \pi_K(P_1(F) \times Q_1(K,L)) = \pi_K Q_1(K,L) \leftrightarrow \square R. \square;$$

$$c) \pi_L(P_1(F) \times Q_1(K,L)) = \pi_L Q_1(K,L) \leftrightarrow \square R^-. \square;$$

$$d) \pi_{K,L}(P_1(F) \times Q_1(K,L)) = \pi_{K,L} Q_1(K,L) \leftrightarrow R;$$

$$e) \pi_{F,K}(P_1(F) \times Q_1(K,L)) = P_1(F) \times \pi_K Q_1(K,L) \leftrightarrow \text{id}(C) \circ \square \square \circ \text{id}(\square R. \square);$$

$$f) \pi_{F,L}(P_1(F) \times Q_1(K,L)) = P_1(F) \times \pi_L Q_1(K,L) \leftrightarrow \text{id}(C) \circ \square \square \circ \text{id}(\square R^-. \square).$$

8.3. *Произведение двух бинарных отношений.* Произведение двух бинарных отношений даёт четырехарное отношение, которое, как и в предыдущем случае, необходимо спроецировать на один или два атрибута, чтобы получить допустимое результирующее отношение. Отметим, что данное отображение следующим образом сводится к предыдущему:

a) проекция на один атрибут сводится к вариантам 8.2.b) или 8.2.c), например:

$$\pi_N(Q_1(K,L) \times Q_2(M,N)) = \pi_N Q_1(M,N) \leftrightarrow \square S^-. \square;$$

b) проекция на два атрибута, оба из которых принадлежат одному отношению, сводится к варианту

8.2.d), например:

$$\pi_{M,N}(Q_1(K,L) \times Q_2(M,N)) = \pi_{M,N}Q_2(M,N) \leftrightarrow S;$$

с) проекция на два атрибута, которые принадлежат разным отношениям, сводится к вариантам 8.2.e) или 8.2.f) , например:

$$\pi_{L,M}(Q_1(K,L) \times Q_2(M,N)) = \pi_L Q_1(K,L) \times \pi_M Q_2(M,N) \leftrightarrow \text{id}(\square R^- . \square) \circ \square \square \circ \text{id}(\square S . \square).$$

Отображение операций RA^2 , которые отсутствуют в RA

1. Инверсное деление

В работе [3] было показано, что операция инверсного деления выражается через совокупность операций классической реляционной алгебры. Основываясь на результатах этой работы, механизмы отображения для операции инверсного деления будут выглядеть следующим образом:

$$Q_1(K,L)[L/F]P_1(F) = \pi_K(Q_1) - \pi_K(Q_1 \square (\pi_K(Q_1) \times (\pi_L(Q_1) - P_1))) \leftrightarrow \\ \square R . \square - \square (R \square (\text{id}(\square R . \square) \circ \square \square \circ \text{id}(\square R^- . \square - C))).$$

2. Номинал

В дескриптивных логиках номинал – это конструктор концепта, который строит концепт из индивида. Если d есть имя индивида, то $\{d\}$ есть концепт. В нашу RA^2 эта операция введена для поддержания конструктора номинала. Поэтому, её отображение не представляет особых проблем.

Отображение операции номинала выглядит следующим образом:

$$\{d\} \leftrightarrow \{d\}.$$

3. Транзитивное замыкание

Эта операция также была включена в RA^2 для поддержания соответствующего конструктора DL . Поэтому её отображение сходно с отображением операции номинала.

Отображение операции транзитивного замыкания выглядит следующим образом:

$$Q_1^+ \leftrightarrow R^+.$$

Выводы

Основная цель данной работы – установление отображений между дескриптивной логикой и реляционной моделью данных. Приведено краткое описание бинарной реляционной модели данных, бинарной реляционной структуры данных, бинарной реляционной алгебры. Рассматриваются механизмы отображения бинарной реляционной модели данных с равенством в дескриптивную логику ALC и её расширения.

Литература

1. Андон Ф.И., Резниченко В.А., Чистякова И.С. Отображение дескриптивной логики в реляционную модель данных. *Кибернетика и системный анализ*. 2017. № 6. С. 160–175.
2. Чистякова И.С. Онтолого-ориентированная интеграция данных в семантическом вебе. *Проблемы програмування*. 2014. № 2–3. С. 188–196.
3. Резниченко В.А., Чистякова И.С. Отображение дескриптивной логики ALC в бинарную реляционную структуру данных. *Проблемы програмування*. 2015. № 4. С. 13–30.
4. Резниченко В.А., Чистякова И.С. Интеграция семейства расширенных дескриптивных логик с реляционной моделью данных. *Проблемы програмування*. 2016. № 2–3. С. 38–47.
5. Чистякова И.С. Интеграция логик с операциями над ролями с реляционной моделью данных. *Проблемы програмування*. 2016. № 4. С. 58–65.
6. Чистякова И.С. Интеграция аксиоматики дескриптивных логик с реляционной моделью данных. *Проблемы програмування*. 2017. № 1. С. 51–58.
7. Резниченко В.А., Чистякова И.С. Бинарная реляционная модель данных. *Проблемы програмування*. 2017. № 2. С. 96–105.
8. 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.
9. Ziegler P., Dittrich K.R. Data Integration – Problems, Approaches, and Perspectives. *Conceptual Modelling in Information Systems Engineering*. 2007. P. 39–58.
10. Richard Barker. “Case*Method: Entity Relationship Modelling”. Addison-Wesley. 1990. 240 p.
11. Пасічник В.В., Резніченко В.А. Організація баз даних та знань. К.: Видавнича група BHV, 2006. 384 с.
12. Золін Е.Е. «Дескрипційна логіка (годовий спецкурс)» [Електронний ресурс]. URL:<http://lpcs.math.msu.su/~zolin/dl/> (дата звернення 01.03.2018).

References

1. Andon, F.I., Reznichenko, V.A., Chystiakova, I.S. (2017) Mapping of description logic to the relational data model. *New tools of cybernetics, informatics, computer engineering, and systems analysis*. 53(6). P. 963–977.
2. Chystiakova, I.S. (2014) Ontology-oriented data integration on the Semantic Web (Онтолого-ориентированная интеграция данных в семантическом вебе). *Problems in programming*. 2-3. P. 188–196.
3. Reznichenko, V.A. & Chystiakova, I.S. (2015) Mapping of the Description Logics ALC into the Binary Relational Data Structure (Отображение дескриптивной логики ALC в бинарную реляционную структуру данных). *Problems in programming*. 4. P. 13–30.
4. Reznichenko, V.A. & Chystiakova, I.S. (2016) Integration of the family of extended description logics to the relational data model (Интеграция семейства расширенных дескриптивных логик с реляционной моделью данных). *Problems in programming*. 2-3. P. 38–47.
5. Chystiakova, I.S. (2016) Integration of the description logics with role extensions to the relational data model (Интеграция логик с операциями над ролями с реляционной моделью данных). *Problems in programming*. 4. P. 58–65.
6. Chystiakova, I.S. (2017) Integration of the description logics axiomatic into relational data model (Интеграция аксиоматики дескриптивных логик с реляционной моделью данных). *Problems in programming*. 1. P. 51–58.
7. Reznichenko, V.A. & Chystiakova, I.S. (2017) Binary relational data model (Бинарная реляционная модель данных). *Problems in programming*. 2. P. 96–105.
8. CODD, E.F. (1979) Extending the database relational model to capture more meaning. *ACM Transactions on Database Systems (TODS)*. 4(4). P. 397–434.
9. Ziegler, P. & Dittrich, K.R. (2007) Data Integration – Problems, Approaches, and Perspectives. *Conceptual Modelling in Information Systems Engineering*. P. 39–58.
10. Barker, R. (1990) *Case*Method: Entity Relationship Modelling*. Addison-Wesley. 240 p.
11. Pasichnyk, V.V. & Reznichenko, V.A. (2006) *Database and Knowledge Base Organization (Організація баз даних та знань)*. BHV. 384 p.
12. Zolin, E.E. (2017) Description logic (one year special course) (Дескрипційна логіка (годової спеціалізації)) [Online]. Available from: <http://lpcs.math.msu.su/~zolin/dl/>

Об авторе:

Чистякова Инна Сергеевна,

младший научный сотрудник Института программных систем НАН Украины.

Количество научных публикаций в украинских изданиях – 10.

Количество научных публикаций в зарубежных изданиях – 1.

orcid.org/0000-0001-7946-3611.

Место работы автора:

Институт программных систем НАН Украины.

03187, г. Киев, проспект Академика Глушкова, 40.

Тел.: +38(066) 847 7784.

E-mail: inna_islyamova@ukr.net.