

**Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка**

ЧИСТЯКОВА Інна Сергіївна

УДК 004.652

**МЕТОД ВІДОБРАЖЕННЯ МІЖ ДЕСКРИПТИВНОЮ ЛОГІКОЮ ТА
РЕЛЯЦІЙНОЮ МОДЕЛЛЮ ДАНИХ**

**01.05.03 – математичне та програмне забезпечення обчислювальних
машин і систем**

**Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук**

Київ – 2021

Дисертацією є рукопис.

Робота виконана в Інституті програмних систем НАН України.

Науковий керівник: кандидат фізико-математичних наук, с.н.с,
Резніченко Валерій Анатолійович,
Інститут програмних систем
НАН України, заступник завідувача відділом №11
автоматизованих інформаційних систем

Офіційні опоненти: доктор технічних наук, професор
Сидоров Микола Олександрович,
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря
Сікорського», кафедра інформатики та
програмної інженерії, професор

кандидат технічних наук, доцент
Франчук Олег Васильович,
Національний університет «Києво-Могилянська
академія», кафедра мережевих технологій,
доцент

Захист відбудеться «09» вересня 2021 р. о(об) 10:00 годині
на засіданні спеціалізованої вченої ради Д 26.001.09 Київського національного
університету імені Тараса Шевченка за адресою:
03127, м. Київ, проспект Академіка Глушкова, 4-д, корп. 6, ф-т кібернетики,
ауд. 40.

З дисертацією можна ознайомитися у Науковій бібліотеці ім. М. Максимовича
Київського національного університету імені Тараса Шевченка за адресою:
01033, м. Київ, вул. Володимирська, 58.

Автореферат розісланий «__» серпня 2021 р.

Учений секретар
спеціалізованої вченої ради

Шевченко В.П.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність проблеми. З початку 70-х років минулого століття широко розвивається один з найбільш затребуваних напрямків сучасної інформаційної індустрії – інтеграція даних. Його розвиток відслідковується таким чином: відбувається формування більш чіткого уявлення про багаторівневу архітектуру баз даних (БД), про моделі даних (МД) як інструмент інформаційного моделювання реальності та про відображення між ними. На першому етапі задача зводилася до підтримки глобальної схеми для сукупності локальних схем, що підтримують одну й ту саму (або різні) МД, які функціонують у різних вузлах під керуванням однієї системи управління базами даних (СУБД). У подальшому задача була узагальнена та включала в себе створення сховищ даних, різноманітних репозиторій інформаційних ресурсів та веб-додатків.

В наш час для роботи з даними інтернет-простору широко застосовуються різноманітні інформаційні системи (ІС). Кожна з них використовує різні засоби зберігання, оперування, маніпулювання, пошуку даних, розмежування доступу до них як всередині так і ззовні ІС. В основі кожної з них лежить своя МД, яка ґрунтується на власному математичному апараті, що є головним джерелом труднощів при взаємодії цих систем. Таким чином, гетерогенність даних, ускладнює процеси використання та зберігання їх в інформаційному просторі. А отже, на сьогоднішній день ключовим питанням інтеграції даних є інтеграція неоднорідних (гетерогенних) даних різноманітних інформаційних систем.

Створення відображень між дескриптивною логікою та реляційною моделлю даних (РМД) є однією з ключових задач проблеми інтеграції даних. Кожного дня об'єм інформації зростає в геометричній прогресії, а її релевантність – в арифметичній. Це породжує низку актуальних питань, пов'язаних із використанням та зберіганням даних інформаційного простору. За різними оцінками, від 75% до 90% ІС використовують реляційні бази даних (РБД) для збереження даних ІС. На решту 10-25% припадає чимала кількість даних, що зберігається в структурах, які визначаються МД, відмінними від реляційної. У цей час продовжує активно розвиватись концепція організації інформаційного простору – семантичний веб. Для роботи з даними створюються онтології, побудовані на основі різноманітних діалектів дескриптивних логік, які в дисертаційній роботі розглядаються у якості самостійних МД. Такий стан справ породжує наступні питання. По-перше, як організувати зберігання онтологій ІС за допомогою реляційних баз даних (РБД). По-друге, яким чином існуючі дані РБД занурити в семантичне середовище та зробити придатними для онтологій. Створення відображень між дескриптивною логікою та РМД покликане дати відповідь на ці питання.

Методології створення відображень між дескриптивною логікою та РМД присвячені роботи І. Астрової, А. Калья, Н. Корда, Л. Стоянович, Н. Стоянович, Р. Вольц, С. Мартінез-Круз, І. Ж. Бланко, М. А. Віла, К. Церанса, Г. Буманса, С. Дас, С. Сундара, Р. Циганяка, С. Р. Упадхая, П. С. Кумар, Е. Виснякас, Л. Немурате та інших.

Робота по дослідженню дескриптивних логік та їх використання у різних прикладних застосуваннях висвітлена у працях Н. Гуріно, П. Гіаретта, Ф. Баадер, Д. Кальванез, Д. МакГінез, Д. Нарді, П. Патель-Шнайдер, І.Ф. Круз, Х. Ксяо, Е.Е. Золіна, А.А. Бездушного, М.О. Сидорова, М.М. Глибовця, С.Л. Кривого, О.В. Франчука, П.І. Андона, В.А. Резніченка, О.В. Новіцького та інших.

В наш час існує декілька специфікацій мов відображення РБД в онтологію, створену за допомогою мов OWL та RDF. До їх складу входить R2R ML (RDB to RDF Mapping Language), що має рекомендацію консорціуму W3C. Слід відзначити мову RDB2OWL, автори якої пропонують власний програмний інструмент для її реалізації.

Як показує огляд літератури, усі існуючі мови відображення мають ряд особливостей, які суттєво звужують їх практичне використання. На сьогоднішній день не існує такого алгоритму, який би реалізовував можливість двостороннього механізму відображення між DL та РМД, зі збереженням повноти та цілісності даних. Одним із підходів є пряме перетворення основних конструкцій DL у РМД та навпаки, з подальшою автоматизацією описаного процесу; також зустрічаються спроби формалізації розглянутого підходу та побудови певної мови відображення (мова RDB2OWL, тощо).

В зв'язку з цим робота є актуальною, оскільки створення механізмів відображення існує лише на рівні структури, але не стосується операцій, що значним чином знижує функціональні можливості систем, які вирішують питання таких перетворень. В свою чергу, дисертаційна робота присвячена вирішенню питання створення відображень, з урахуванням операційної складової.

Зв'язок роботи з науковими програмами, планами, темами. Робота виконувалася відповідно до планів наукових досліджень відділу автоматизованих інформаційних систем Інституту програмних систем НАН України в рамках науково-дослідних тем: «Розробка методів та засобів підтримки побудови інтелектуальних інформаційних систем у семантичному Веб-середовищі», шифр III-2-12, № держреєстрації 0112U002761, (2012-2016); «Інтелектуальні інформаційні системи підтримки аналізу великих даних (Big Data)», шифр III-10-18, № держреєстрації 0118U001906, (2018-2019); «Методи та засоби створення інтелектуальних сервіс-орієнтованих інформаційно-забезпечуючих систем у середовищі семантичного Вебу», шифр III-2-17, № держреєстрації 0117U000736, (2017-2021).

Мета і задачі дослідження. Мета дисертаційної роботи – розробка, обґрунтування та апробація відображень між дескриптивною логікою та реляційною моделлю даних.

Задачі дослідження:

- визначити та створити схему інтеграції даних;
- розробити концептуальну інформаційну ER-модель дескриптивної логіки;
- створити бінарну реляційну модель даних (RM²);

- розробити двосторонні механізми відображення між попередньо визначеними діалектами DL та RM^2 ;

- здійснити апробацію розробленого двостороннього механізму відображень за допомогою RDF.

Об'єкт дослідження – проблема інтеграції даних в семантичному вебi.

Предмет дослідження – засоби та технологія відображення між дескриптивною логікою та реляційною моделлю даних.

Методи дослідження. Вирішення завдань, поставлених в дисертаційній роботі, виконано з використанням дескриптивних логік, математичної логіки, методології концептуального інформаційного моделювання предметних областей, теорії нормалізації реляційної моделі даних, методів проектування реляційних баз даних.

Наукова новизна отриманих результатів.

В роботі отримано нові наукові результати, які виносяться на захист:

- побудовано таксономію досліджень, які стосуються відображень між онтологіями та реляційними базами даних; визначено вузькі місця напрямків досліджень, з яких складається визначена таксономія; обґрунтовано важливість опису відображень між дескриптивною логікою та реляційною моделлю даних, а також вибір централізованої схеми інтеграції даних ІС, керуючись результатами проведеного аналізу проблеми інтеграції даних в семантичному вебi;

- уперше побудовано концептуальну інформаційну ER-модель дескриптивної логіки із залученням технології, яка забезпечує: підтримку гіпотези відкритого світу та не унікальності імен, повне представлення усіх інформаційних складових дескриптивної логіки, використання тільки унарних та бінарних сутностей, безпосереднє її відображення у бінарну реляційну структуру даних зі схемою бази даних у третій нормальній формі;

- визначено *спеціальний* варіант класичної РМД – бінарну реляційну модель даних (RM^2), яка складається з бінарної реляційної структури даних (RS^2), бінарної реляційної алгебри (RA^2), та обмежень цілісності. Оскільки, будь-яке n -арне відношення можна представити сукупністю бінарних, було запропоновано спосіб перетворення n -арної (класичної) реляційної структури в бінарну;

- уперше визначено бінарну реляційну алгебру (RA^2), до якої внесено принципово нові операції по відношенню до класичної реляційної алгебри (RA), для підтримки відображення конструкторів концептів та ролей DL. Операції RA , які не збільшують арність результуючого відношення внесено в RA^2 без змін. Операції RA , що утворюють відношення арністю більше 2, були внесені в RA^2 у модифікованому вигляді, такому, що результат їх виконання теж утворює унарне або бінарне відношення;

- уперше розроблено метод двостороннього відображення між онтологічною та реляційною моделями даних, який: носить формальний характер; визначається як на структурному, так і на операційному рівні (включено конструктори концептів та ролей, а також операції реляційної алгебри); передбачає відображення термінологічних аксіом дескриптивної логіки;

- проведено програмну апробацію алгоритму відображення за допомогою мови RDF на прикладі тестових даних; визначено алгоритмічну оцінку складності.

Практичне значення одержаних результатів. Розроблений метод відображення було використано в Інституті розробки інформаційних систем з метою підвищення інтелектуальності автоматизованих інформаційних систем шляхом перетворення реляційних інформаційних ресурсів в онтологічну базу даних (акт упровадження від 7 червня 2021р).

Метод відображення було впроваджено у компанії Зініт Солюшенс Юкрейн під час проектування автоматизованої системи комплектації обчислювальної техніки (акт упровадження від 3 червня 2021р).

Алгоритм відображення між дескриптивною логікою та реляційною моделлю даних може бути використано для створення бібліотеки функцій при розробці програмного забезпечення що підтримує відображення між існуючими онтологіями та базами даних.

Особистий внесок здобувача. Усі наукові результати та положення, які складають основну частину дисертаційної роботи, отримано здобувачем самостійно. Визначення напрямку досліджень та загальна постановка задачі належить науковому керівнику. Роботи [2, 3, 5, 6, 8, 9, 10, 11, 13] написано здобувачем особисто і всі наукові результати, представлені в них, є оригінальними. У працях, опублікованих у співавторстві [1, 4, 7, 12], дисертанту належать наступні результати: [1] – побудовано бінарну реляційну структуру даних, розроблено та описано алгоритм відображення для базових складових дескриптивної логіки ALC, а також обраних розширень та термінологічних аксіом; [4] – побудовано концептуальну інформаційну модель дескриптивної логіки; [7] – створено та описано бінарну реляційну модель даних та її складові частини; [12] – розроблено та описано алгоритм відображення для обраних чисельних розширень ALC та номіналу.

Апробація результатів дисертації. Основні ідеї, принципи, положення та результати доповідалися: на Міжнародній науково-практичній конференції з програмування «УкрПрог 2014» (м. Київ, 20-22 травня 2014); на Міжнародній науково-практичній конференції з програмування «УкрПрог 2016», (м. Київ, 24–25 травня 2016); на Міжнародній науково-практичній конференції з програмування «УкрПрог 2018», (м. Київ, 24–25 травня 2018); на міжнародній науковій конференції студентів та молодих науковців «Aviation in the XXI-st century» (м.Київ, 21–23 листопада 2014); на ІХ Міжнародній науково-практичній Інтернет конференції «Глобальні та регіональні проблеми інформатизації в суспільстві та природокористуванні 2021» (м.Київ 13-14 травня 2021).

Публікації. Основні наукові результати дисертації опубліковано в 14 наукових працях, із них 9 наукових – у виданнях з переліку наукових фахових видань України, з яких одна наукова робота [1] у міжнародному науково-технічному журналі, що перекладається у США англійською мовою у видавництві Springer під назвою «Cybernetics and Systems Analysis» та індексується в базі SCOPUS, 5 тез [10–14] опубліковано у матеріалах

міжнародних наукових конференцій, з яких матеріали [10, 12, 13] індексуються в SCOPUS.

Структура та обсяг дисертації. Дисертаційна робота складається із вступу, чотирьох розділів, загальних висновків, списку використаних джерел із 99 найменувань. Загальний обсяг дисертації становить 174 сторінки, у тому числі основний текст викладено на 115 сторінках. Робота містить 13 рисунків, 12 таблиць та 7 додатків.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми дисертаційної роботи, визначено мету та об'єкт, предмет дослідження, сформульовано основні задачі роботи, зазначено характеристику наукової новизни та практичної значимості отриманих результатів. Визначено особистий внесок здобувача у кожній опублікованій науковій праці.

Перший розділ присвячено огляду літературних джерел, пов'язаних з тематикою дисертації та обґрунтуванню вибору напрямку досліджень.

Розглянуто причини виникнення гетерогенності даних. Наведено огляд проблеми інтеграції даних в семантичному веб, а саме – три її складові частини: схеми інтеграції даних, відображення між МД, засоби маніпулювання даними. Наведено огляд схем інтеграції даних та обрано централізовану схему у якості інтегруючої. Дано характеристику двох різновидів централізованої архітектури: Global-as-View (GAV) та Local-as-View (LAV). Ґрунтуючись на цьому, дано визначення семантичної інтеграції даних.

Семантична (онтолого-орієнтована) інтеграція даних – це використання онтології у якості об'єднуючої (глобальної) моделі для опису і підтримки відображень між різними МД, а також уніфікованого маніпулювання даними.

У випадку використання централізованої схеми інтеграції даних, для вирішення комплексної проблеми інтеграції, онтологія найкраще підходить для глобальної моделі, що дозволяє у якості локальної моделі застосовувати будь-яку МД. У зв'язку з цим, у якості центральної моделі інтеграційної схеми пропонується обрати дескриптивну логіку (DL).

Наведено аргументацію твердження **DL можна розглядати як модель даних**:

1. DL характеризується синтаксисом та семантикою. Саме у семантиці проявляється її структурна частина (виникають поняття множин, бінарних відношень і т. д.).

2. DL характеризується маніпулятивною частиною, яку представляють такі її компоненти, як конструктори концептів та ролей.

3. DL характеризується цілісною частиною, яку представляють такі її компоненти, як аксіоми.

У якості локальної моделі обрано реляційну модель даних (РМД).

Наведено огляд досліджень по встановленню відображень між онтологічною та реляційною моделями даних. Створено класифікацію

досліджень та розробок в області опису відображень між зазначеними моделями даних. Проаналізовано сукупність досліджень кожного із напрямків таксономії. Визначено та узагальнено вузькі місця, притаманні кожному з напрямків.

Класифікація представлена у вигляді таксономії, яка представлена на рис. 1.

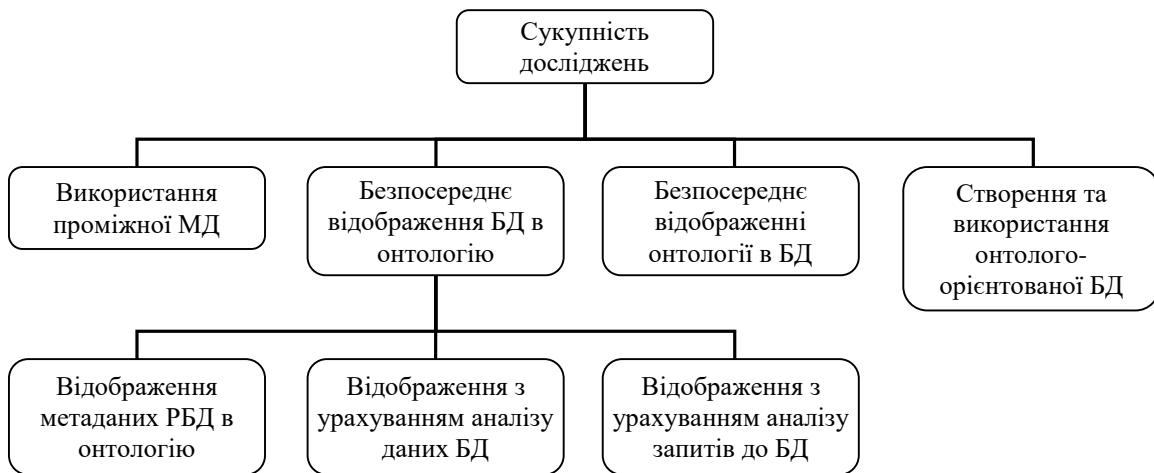


Рис. 1. Таксономія досліджень підходів до створення відображень між РМД та онтологіями

У другому розділі визначено підмножину реляційної моделі даних та дескриптивної логіки, які приймають участь у створенні відображень; обґрунтовано необхідність створення спеціального варіанту РМД, а саме – бінарної реляційної моделі даних (RM^2); показано спосіб перетворення n-арної реляційної структури в бінарну; визначено базову дескриптивну логіку ALC, перелік її розширень та перелік аксіом, для яких створено відображення; розглянуто механізм реіфікації дескриптивних логік.

У підрозділі 2.1 визначено бінарну реляційну модель даних та її основні складові.

Бінарна реляційна модель даних (RM^2) – це сукупність бінарної реляційної структури даних, бінарної реляційної алгебри та обмежень цілісності.

RM^2 включає в себе:

1. *Структурний аспект*, що представлений бінарною реляційною структурою даних, у якій дані представлені у вигляді сукупності відношень;
2. *Аспект маніпулювання*, що представлений бінарною реляційною алгеброю, яка включає в себе оператори маніпулювання відношеннями;
3. *Аспект цілісності*, що представлений обмеженнями цілісності, яким відповідають відношення бінарної реляційної структури даних.

Бінарна реляційна структура даних (RS^2) – це сукупність реляційних відношень арності не більше 2.

Бінарна реляційна алгебра (RA^2) – це сукупність операцій, що визначені на бінарній реляційній структурі. Вона замкнена у тому сенсі, що результат виконання будь-якої операції теж є відношенням.

Існують три типи операцій:

- *ті, що не збільшують арність результуючого відношення.* Вони є повними аналогами операцій класичної реляційної алгебри (RA).

- *ті, що збільшують арність результуючого відношення.* Вони мають свої аналоги в RA. Враховуючи це, в RA^2 вводяться їх модифікації, результуючі відношення яких не перевищують арності 2.

- *ті, що є новими по відношенню до RA.*

До **першого типу** операцій відносяться ті операції, що не збільшують арність результуючого відношення, а саме – іменування, об'єднання, перетин, різниця, проекція, селекція, ділення.

До **другого типу** операцій відносяться операції декартового добутку та з'єднання. Оскільки у результаті виконання цих операцій збільшується арність результуючого відношення, у RA^2 вводяться модифіковані варіанти цих операцій, результуючі відношення яких не перевищують арності 2. Назви операцій відповідно збережені.

До **третього типу** операцій відносяться: інверсне ділення, номінал, транзитивне замикання, рефлексивно-транзитивне замикання.

Операція **інверсного ділення (/)** наступним чином виражається через уже визначені оператори:

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

де $R_1(A, B)$ та $R_2(C, D)$ – відношення, причому B та C – порівнянні за предикатом рівності.

Суть операції **номіналу ({})** – побудова унарного чи бінарного відношення із заданої константи або пари констант. В RA така операція відсутня, тому її було включено для підтримки однойменного конструктора DL. Крім того, усі реляційні СУБД підтримують цю операцію.

Операція **транзитивного замикання(+)** відсутня в RA, однак ця проблема вирішена в реляційних СУБД, що підтримують рекурсивний SQL.

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

де $R(A, B)$ – бінарне відношення.

Операція **рефлексивно-транзитивного замикання(*)** очевидним чином виражається через операції проекції, з'єднання, транзитивного замикання та об'єднання.

$$R^* = (\pi_A([A, B]) \bowtie_{A=A} (\pi_A([A, B])) \cup R^+.$$

Цілісність даних – (в класичній РМД) це механізм підтримки відповідності структури даних предметної області.

Обмеження цілісності – це правила, що накладають обмеження на можливий стан структури даних.

В бінарній реляційній моделі даних, аналогічно класичній РМД, визначено дві базових вимоги забезпечення цілісності: цілісність сутностей та цілісність зв'язків. Як правило, підтримка цілісності сутностей та зв'язків забезпечується засобами системи управління базою даних (СУБД).

У **підрозділі 2.2** описано алгоритм перетворення n-арної реляційної структури в бінарну. Класична РМД не обмежується лише унарними та бінарними відношеннями. Доведено, що будь-яке n-арне відношення можна

представити сукупністю бінарних. Спосіб такого перетворення полягає в тому, що воно повинно бути еквівалентним по даним, тобто сукупність відношень R_1, R_2, \dots, R_k є еквівалентною по даним відношення R , якщо R дорівнює їх природному з'єднанню:

$$R = R_1 * R_2 * \dots * R_k.$$

Згідно теоремі Хіта, якщо у відношенні $R(A, B, C)$, де A, B, C у загальному випадку можуть бути наборами атрибутів, існує функціональна залежність $A \rightarrow B$, то $R = R[A, B] * R[A, C]$.

У результаті, якщо атрибут A_1 відношення $R(A_1, A_2, \dots, A_n)$ є первинним ключем, тобто він функціонально визначає всі інші атрибути, то R можна декомпонувати на наступну сукупність бінарних відношень:

$$R[A_1, A_2], R[A_1, A_3], \dots, R[A_1, A_n],$$

причому

$$R = R[A_1, A_2] * R[A_1, A_3] * \dots * R[A_1, A_n].$$

Тобто, така декомпозиція є еквівалентною за даними. Таким чином, якщо у відношенні існує простий (не складений) первинний ключ, то з наведеного вище показано, як можна проводити декомпозицію n -арних відношень у сукупність бінарних.

Решту підрозділу присвячено випадку, коли первинний ключ є складеним. В такій ситуації пропонується використати концепцію простого ключа-замінника для вирішення задачі декомпозиції у бінарні відношення.

У **підрозділі 2.3** визначено підмножину дескриптивної логіки, її розширень конструкторами концептів та ролей, а також перелік аксіом, для яких створено відображення.

Серед усіх сімейств дескриптивних логік було обрано DL ALC. Сучасні DL являють собою багаточисельні розширення логіки ALC, шляхом додавання нових конструкторів для побудови складених концептів та ролей, або додавання нових видів аксіом. Вони розширюють базовий синтаксис саме DL ALC. Тобто, кожна дескриптивна логіка містить у собі повністю весь склад цієї логіки, а також від однієї до декількох додаткових операцій, якими розширюється ALC. Існує неформальна угода за якою розширена логіка отримує своє ім'я: зазвичай, кожному конструктору або виду аксіом привласнюють букву, яка (за деяким виключенням) додається до імені логіки. Таким чином, було утворено більшість сучасних DL, наприклад, ALCIQ, SHIQ, SHOIN, SROIQ і т.д. Це обґрунтовує першочерговий вибір ALC для створення відображень між DL та РМД.

Коротка форма синтаксису ALC:

$$T \mid \perp \mid C \mid R \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \exists R.C \mid \forall R.C$$

У дисертаційній роботі створено відображення для таких розширень ALC: чисельні обмеження на ролі (функціональні, кількісні та якісні), номінал, конструктори ролей (обернена роль, доповнення ролі, перетин та об'єднання ролей, композиція, ролефікація, транзитивне та рефлексивно-транзитивне замикання).

Концепти DL використовуються як інструмент запису знань про предметну область. Вони мають інтенціональну та екстенціональну складову.

Інтенціональні – загальні знання про поняття та їх взаємозв'язки.

Екстенціональні – знання про індивідуальні об'єкти, їх властивості та зв'язки з іншими об'єктами.

У зв'язку із цим розподілом, знання, що записуються за допомогою DL, підрозділяються на: TBox(T) - набір термінологічних аксіом; ABox (A) – набір тверджень (фактів) про індивіди; RBox (R) – набір аксіом для ролей, який є спеціальним розширенням синтаксису логіки ALC.

Термінологічною аксіомою називається вираз вигляду $C \sqsubseteq D$ або $C \equiv D$, де C і D – довільні концепти. Термінологією (TBox) називається довільний кінцевий набір аксіом даного вигляду.

Системою фактів (ABox) називають кінцеву множину A тверджень вигляду $a:C$ та aRb , де a, b – індивіди, C – довільний концепт, R – роль.

Для системи фактів іноді вводиться «угода про унікальність імен», яка означає, що різним іменам індивідів відповідають різні елементи області інтерпретації. Якщо угоду не висунуто (тобто допускається, що один об'єкт може мати багато імен), тоді вводиться **аксіома рівності індивідів**. За допомогою цієї аксіоми можна вказати, що певна пара імен іменує один індивід, тобто **не виконується принцип унікальності**. Однак, одне ім'я іменує один індивід, як і раніше.

Аксіоми ролей розширюють DL ALC так само, як і додавання до неї нових конструкторів концептів та ролей. Існує чотири аксіоми ролей: H – ієрархія ролей; S – транзитивні ролі; \equiv – еквівалентні ролі; \neq – різні ролі.

DL ALC та більшість сучасних дескриптивних логік містять у собі лише 1-місні та 2-місні предикати – концепти та ролі відповідно. На практиці інколи виникає потреба у мові, що представляє знання про багатомісні відношення. З цією метою різними авторами було запропоновано декілька розширень DL. Найбільш відоме з них – логіка DLR, яка зводиться до відомої логіки ALCIQ, за допомогою техніки реіфікації (або матеріалізації).

Третій розділ присвячено створенню відображень між дескриптивною логікою та бінарною реляційною моделлю даних.

У загальному випадку, відображення M (відображення між глобальною схемою, та схемою джерела) уявляє собою набір тверджень у вигляді $q_S \rightarrow q_G$ та $q_G \rightarrow q_S$, де q_G та q_S – сумісні запити, які сформульовані в термінах моделей глобальної та локальної схем відповідно. Окремим випадком є ситуація, де замість запитів використовуються структурні елементи відповідних моделей. Тобто, відображення DL у RM^2 уявляє собою набір тверджень, по одному на кожний елемент із DL, пов'язаних запитом із відповідним набором елементів RM^2 . У свою чергу, відображення RM^2 у DL уявляє собою набір тверджень, по одному на кожний елемент RM^2 , пов'язаних запитом із відповідним набором елементів DL.

У **підрозділі 3.1** запропоновано схему відображення з проміжною побудовою концептуальної інформаційної моделі DL. В дисертаційній роботі не ставилася задача представити механізм ризонерів дескриптивної логіки в реляційній моделі даних. Взаємне відображення відбувається лише між двома

моделями, включно зі структурою та операціями. Схема концептуальної інформаційної моделі дескриптивної логіки показана на рис.2.

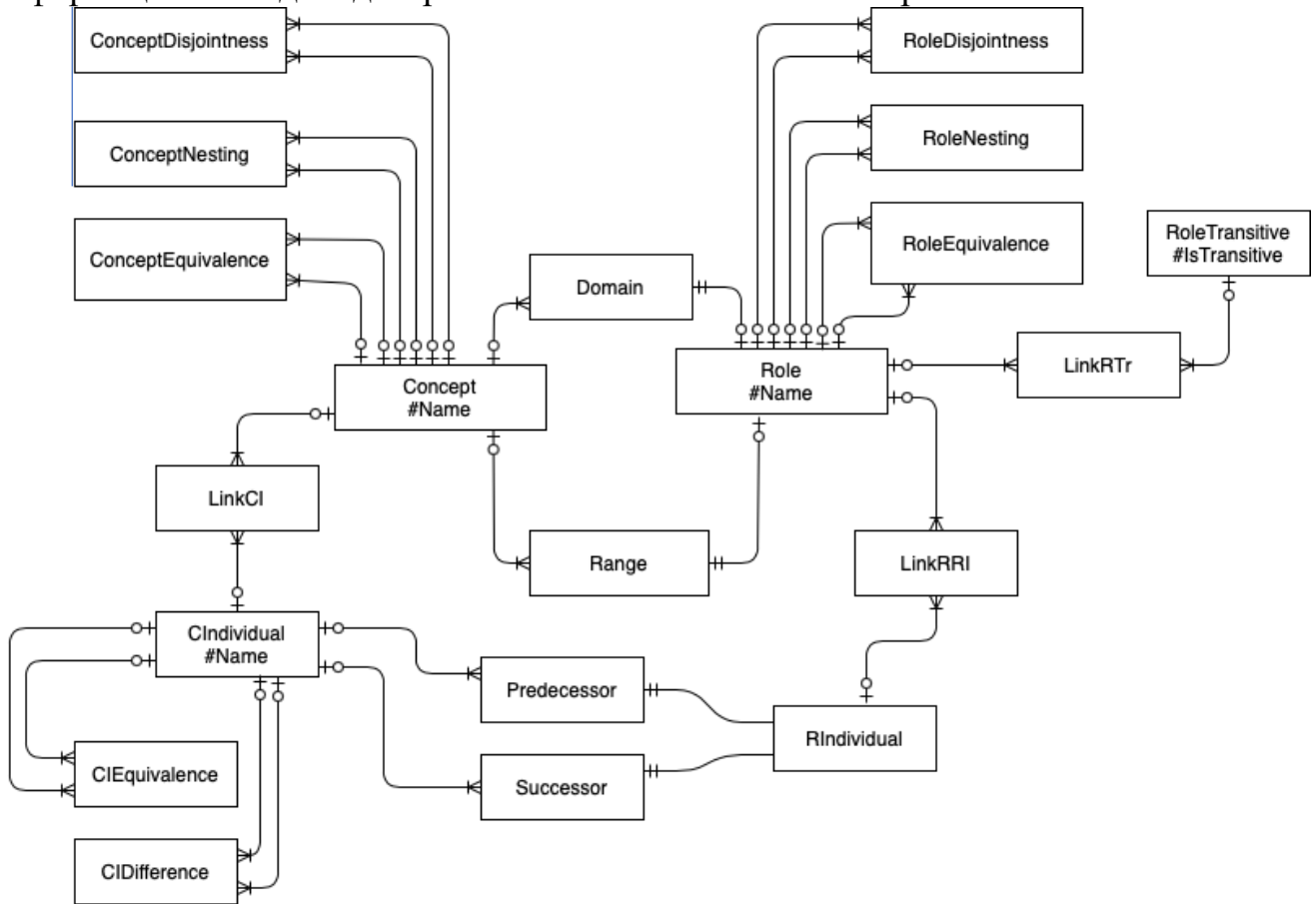


Рис. 2. Концептуальна інформаційна модель дескриптивної логіки.

Концептуальну ER-схема побудовано з використанням наступних методологічних правил:

1) Імена індивідів концептів та ролей, а також імена самих концептів та ролей представлено самостійними незалежними унарними сутностями. Такі сутності мають єдиний атрибут "Ім'я", що унікально їх ідентифікує. Незалежність цих сутностей визначається тим, що всі закінчення їх зв'язків факультативні (тобто екземпляри таких сутностей існують незалежно від їх можливого зв'язку з екземплярами інших сутностей).

2) Зв'язок між парою сутностей (п.1) реалізовано за допомогою введення додаткової перехідної сутності, яка не має власних атрибутів. Обидва закінчення зв'язків перехідної сутності мають тип "багато". Ці зв'язки обов'язкові і входять у склад унікального ідентифікатора. Якщо між парою сутностей (п.1) багато зв'язків, то для кожної з них будується перехідна сутність.

3) Якщо та чи інша сутність має власний атрибут, то для забезпечення бінарності створюється дві додаткових сутності. Перша – це самостійна унарна сутність з єдиним ідентифікуючим атрибутом, що успадковує атрибут початкової сутності. Друга – це перехідна сутність для встановлення зв'язку між створеною унарною сутністю, та початковою.

Концептуальна інформаційна ER-схема DL, що побудована за цими правилами, надає можливість підтримувати наступні властивості гіпотези відкритого світу та не унікальності імен:

- кожен індивід концепту/ролі може існувати самостійно, тобто незалежно від самих концептів/ролей;
- індивіди концептів/ролей можуть мати властивості, що не характерні для концептів/ролей, яким вони належать;
- індивіди можуть мати декілька імен.

Побудовано схему RM^2 , яка відповідає наведеній концептуальній інформаційній моделі. Оскільки ER-схема відповідає вимогам 3NF, то кожна сутність представляється у вигляді відношення RM^2 . Наведено усі сутності ER-моделі з описом та відповідні відношення з переліком атрибутів.

У **підрозділі 3.2** представлено спосіб відображень DL ALC у RM^2 . Для кожного конструктора концептів та ролей наведено покроковий алгоритм утворення формули відображення, що уявляє собою сукупність операцій RA^2 . Наведено формули для відображення: концепту, ролі, доповнення концепту, об'єднання концептів, перетину концептів, конструктору екзистенціального обмеження.

Відображення конструктора обмеження значення. У результаті відображення буде отримано множину таких елементів, усі R-наступники яких будуть належати інтерпретуючій множині концепту C.

Конструктор обмеження значення має наступну семантику:

$$\forall y(Q(x, y) \supset P(y))$$

по заданій **множині** знайти всі її **підмножини**, які задовольняють заданій властивості.

Відображення для конструктора обмеження значення використовує операцію інверсного ділення і виглядає наступним чином:

$$(\forall R. C)_{RM^2}^E = \pi_{First}(R_{RM^2}^E [Second/Name] C_{RM^2}^E)$$

Такий варіант визначається через інші операції RA^2 наступним чином:

$$(\forall R. C)_{RM^2}^E = \pi_{First} R_{RM^2}^E - \pi_{First}(R_{RM^2}^E \cap (\pi_{First} R_{RM^2}^E \times (\pi_{Second} R_{RM^2}^E - \pi_{Name} C_{RM^2}^E)))$$

У **підрозділі 3.3** представлено спосіб відображення чисельних розширень ALC, номіналу та групи конструкторів ролей. Для кожного конструктора концептів та ролей наведено покроковий алгоритм утворення формули відображення, що уявляє собою сукупність операцій RA^2 .

В роботі використовується наступний розподіл чисельних обмежень на групи: обмеження функціональності (познач. літерою F); кількісне обмеження (познач. літерою N); якісне обмеження (познач. літерою Q).

Відображення функціонального обмеження (F)

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

Відображення кількісного обмеження (N)

$$(\geq nR)_{RM^2}^E = \pi_{First}(\sigma_{\substack{1 \leq i < j \leq n \\ R_i.Second_i \neq R_j.Second_j}}(\rho_{R_i(First, Second_i)}^*(R_{RM^2}^E)))$$

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

$$\begin{aligned}
(= nR)_{RM^2}^E &= (\geq nR)_{RM^2}^E \cap (\leq nR)_{RM^2}^E \\
(> nR)_{RM^2}^E &= (\geq nR)_{RM^2}^E - (= nR)_{RM^2}^E \\
(< nR)_{RM^2}^E &= (\leq nR)_{RM^2}^E - (= nR)_{RM^2}^E
\end{aligned}$$

Відображення якісного обмеження (Q)

$$(\geq nR.C)_{RM^2}^E = \pi_{\text{First}} \left(\sigma_{1 \leq i < j \leq n} R_i \cdot \text{Second}_i \neq R_j \cdot \text{Second}_j \right) \left(\begin{array}{c} * \\ \text{Second} \end{array} \right)$$

$$\rho_{R_i(\text{First}, \text{Second}_i)} (R_{RM^2}^E \bowtie_{\text{Second}=\text{Name}} C_{RM^2}^2))$$

$$(\leq nR.C)_{RM^2}^E = \pi_{\text{Name}}(C_{\text{Individual}}) - \pi_{\text{First}} \left(\sigma_{1 \leq i < j \leq n} R_i \cdot \text{Second}_i \neq R_j \cdot \text{Second}_j \right) \left(\begin{array}{c} * \\ \text{Second} \end{array} \right)$$

$$\rho_{R_i(\text{First}, \text{Second}_i)} (R_{RM^2}^E \bowtie_{\text{Second}=\text{Name}} C_{RM^2}^E))$$

$$(= nR.C)_{RM^2}^E = (\geq nR.C)_{RM^2}^E \cap (\leq nR.C)_{RM^2}^E$$

$$(> nR.C)_{RM^2}^E = (\geq nR.C)_{RM^2}^E - (= nR.C)_{RM^2}^E$$

$$(< nR.C)_{RM^2}^E = (\leq nR.C)_{RM^2}^E - (= nR.C)_{RM^2}^E$$

Номінал – це конструктор концепту, який будує концепт із індивіда. Якщо d – ім'я концепту, то $\{d\}$ – концепт. Таким чином, якщо ім'я індивіда d інтерпретувалось як $d^I \in \Delta$, то номінал $\{d\}$ інтерпретується як одноелементна множина $\{d^I\} \subseteq \Delta$.

Відображення номіналу виглядає таким чином:

$$\{a\}_{RM^2}^E = \{a_{RM^2}^E\}$$

У роботі представлено спосіб відображень для розширень, що являють собою групу конструкторів ролей, а саме – обернена роль, доповнення ролі, перетин ролей, об'єднання ролей, композиція ролей та ін.

Ролефікація – операція, яка створює нові екземпляри ролей із індивідів заданого концепту.

Відображення ролефікації виглядає наступним чином:

$$(\text{id}(C))_{RM^2}^E = (\rho_{\text{Name}/\text{First}}(\pi_{\text{Name}} C_{RM^2}^E)) \bowtie_{\text{First}=\text{Second}} (\rho_{\text{Name}/\text{Second}}(\pi_{\text{Name}} C_{RM^2}^E))$$

Відображення операції транзитивного замикання

$$(R^+)_{RM^2}^E = (R_{RM^2}^E)^+$$

Відображення операції рефлексивно-транзитивного замикання

Рефлексивно-транзитивне замикання можна виразити через операцію транзитивного замикання:

$$R^* \equiv \text{id}(T) \sqcup R^+$$

Відображення для цього конструктора виглядає наступним чином:

$$(R^*)_{RM^2}^E = (\rho_{\text{Name}/\text{First}}(\pi_{\text{Name}} C_{\text{Individual}})) \bowtie_{\text{First}=\text{Second}} (\rho_{\text{Name}/\text{Second}}(\pi_{\text{Name}} C_{\text{Individual}})) \cup (R^+)_{RM^2}^E$$

У **підрозділі 3.4** представлено відображення термінологічних аксіом дескриптивної логіки. Концептуальна інформаційна модель DL будувалася таким чином, що кожна різновидність термінологічних аксіом концептів, ролей та індивідів була представлена окремою сутністю. Ці сутності при відображенні були представлені окремими реляційними відношеннями. В кожному із

відношень присутні два зовнішніх ключі, що посилаються на концепти, ролі чи індивіди, відносно яких формулюється аксіома.

Синтаксис, семантика термінологічних аксіом DL, а також відповідне відношення в RM^2 подано в таблиці 1.

Таблиця 1— Відображення термінологічних аксіом DL в RM^2

Синтаксис	Семантика	Реляційне відношення	Назва аксіоми
$C \sqsubseteq D$	$C^I \subseteq D^I$	ConceptNesting(CInFK, COutFK)	Вкладеність концептів
$C \equiv D$	$C^I = D^I$	ConceptEquivalence(CForFK, CIsFK)	Еквівалентність концептів
$C_i \sqcap C_j$	$C_i^I \cap C_j^I = \emptyset$ $i \neq j$	ConceptDisjointness(CForDisFK, CIsDisFK)	Неперетин концептів
$R \sqsubseteq S$	$R^I \subseteq S^I$	RoleNesting(RInFK, ROutFK)	Ієрархія ролей
$R \equiv S$	$R^I = S^I$	RoleEquivalence(RForFK, RIsFK)	Еквівалентність ролі
$a = b$	$a^I = b^I$	CIEquivalence(CIForFK, CIIsFK)	Еквівалентність індивідів
$a \neq b$	$a^I \neq b^I$	CIDifferent(CIForDifFK, CIIsDifFK)	Різність індивідів
$Tr(R)$	R^I	RoleTransitive(IsTransitive)	Транзитивна роль

Підрозділ 3.5 присвячено відображенню бінарної реляційної моделі даних в дескриптивну логіку. Наведено правила еквівалентного перетворення для операцій RA^2 , які використовуються при створенні відображень. Для кожної операції бінарної реляційної алгебри наведено покроковий механізм утворення формули відображення, що уявляє собою сукупність конструкторів концептів та ролей DL.

Нехай задані наступні унарні та бінарні відношення:

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

Реляційне відношення має інтенціональну та екстенціональну складову.

Відображення **унарного відношення** в дескриптивну логіку відбувається виходячи з наступних правил: ім'я відношення відображається в ім'я **концепту**; екстенціонал відношення відображається в **інтерпретацію концепту**.

Відображення **бінарного відношення** в дескриптивну логіку відбувається виходячи з наступних правил: ім'я відношення відображається в ім'я **ролі**; -екстенціонал відношення відображається в **інтерпретацію ролі**.

Відображення між атомарними відношеннями, концептами (C, D) та ролями (R, S) задається наступним чином:

$$\begin{aligned} P_1 &\leftrightarrow C, \text{ якщо } P_1^E = C^I \\ P_2 &\leftrightarrow D, \text{ якщо } P_2^E = D^I \\ Q_1 &\leftrightarrow R, \text{ якщо } Q_1^E = R^I \\ Q_2 &\leftrightarrow S, \text{ якщо } Q_2^E = S^I \end{aligned}$$

В роботі наведено відображення для операцій RA^2 , таких як об'єднання, перетин, різниця, проекція, ділення та ін. Причому, операція проекція існує тільки для *бінарних* відношень.

Операція селекції

Операція розглядається тільки за умовою рівності ($=$), оскільки в дисертаційній роботі використовуються лише такі DL, для яких існує тільки умова рівності концептів, ролей та індивідів.

Унарне відношення відображається в концепт DL. В DL не існує операції створення концепту із концепту, але існує операція створення ролі із концепту – $id(C)$. Таким чином, відображення операції селекції для *унарного* відношення відбувається наступним чином: із концепту, що є інтерпретацією унарного відношення, створюється роль за допомогою $id(C)$, а далі виконується селекція за заданою умовою рівності та подальша проекція на перший атрибут.

Відображення операції селекції *унарного* відношення за умовою рівності атрибутів константі виглядає наступним чином:

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

Відображення селекції *бінарного* відношення за умовою рівності атрибутів має вид:

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

Відображення операції селекції, *бінарного* відношення за умовою рівності атрибутів константі описується так:

$$\begin{aligned} \sigma_{L=l}Q_1(K, L) &\leftrightarrow id(\exists R. \{l\}) \circ U \circ id(\{l\}) \\ \sigma_{K=k}Q_1(K, L) &\leftrightarrow id(\exists R^-. \{k\}) \circ U \circ id(\{k\}) \end{aligned}$$

Операція з'єднання

З'єднання двох унарних відношень у результаті дають бінарне відношення

$$P_1(F) \bowtie_{F=G} P_2(G) \leftrightarrow id(C \sqcap D)$$

З'єднання унарного та бінарного відношення дають у результаті трьохарне відношення. Оскільки в RA^2 допустимі тільки бінарні відношення, то необхідно виконати проекцію на один або два атрибути.

$$\pi_F(P_1(F) \bowtie_{F=K} Q_1(K, L)) = \pi_F(P_1(F) \bowtie_{F=K} \pi_K Q_1(K, L)) \leftrightarrow C \sqcap \exists R. T \quad (2)$$

$$\pi_K(P_1(F) \bowtie_{F=K} Q_1(K, L)) = \pi_K(P_1(F) \bowtie_{F=K} \pi_K Q_1(K, L)) \leftrightarrow C \sqcap \exists R. T \quad (3)$$

$$\pi_L(P_1(F) \bowtie_{F=K} Q_1(K, L)) \leftrightarrow \exists R^-. (C \sqcap \exists R. T) \quad (4)$$

$$\pi_{F,K}(P_1(F) \bowtie_{F=K} Q_1(K, L)) \leftrightarrow id(C \sqcap \exists R. T) \quad (5)$$

$$\pi_{K,L}(P_1(F) \bowtie_{F=K} Q_1(K, L)) \leftrightarrow R - (id(\exists R. T - C) \circ U \circ id(\exists R^-. T)) \quad (6)$$

$$\pi_{F,L}(P_1(F) \bowtie_{F=K} Q_1(K, L)) - \text{рівносильно (6)}$$

З'єднання двох бінарних відношень дає чотирьохарне відношення, яке також проектується на один або два атрибути, аналогічно попередньому варіанту.

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

Якщо у результуючій проекції немає атрибуту K, то замінюємо $Q_1(K,L)$ на $\pi_L Q_1(K, L)$ і все зводиться до *з'єднання унарного та бінарного відношень*.

Якщо у результуючій проекції є атрибут K, то отримуємо наступні варіанти:

$\pi_K(Q_1(K, L) \bowtie_{L=M} Q_2(M, N))$ – замінюємо $Q_2(M,N)$ на $\pi_M Q_2(M, N)$ і тим самим зводимо до варіанту (4);

$\pi_{K,L}(Q_1(K, L) \bowtie_{L=M} Q_2(M, N))$ – замінюємо $Q_2(M, N)$ на $\pi_M(M, N)$ і тим самим зводимо до варіанту (6);

$\pi_{K,M}(Q_1(K, L) \bowtie_{L=M} Q_2(M, N))$ – зводиться до варіанту (3)

$\pi_{K,N}(Q_1(K, L) \bowtie_{L=M} Q_2(M, N)) \leftrightarrow R \circ S$

Операція декартового добутку

Відображення для добутку двох унарних відношень наступне:

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

Добуток унарного та бінарного відношень необхідно спроекувати на один чи два атрибути, щоб отримати допустиме результуюче відношення:

$$\pi_F(P_1(F) \times Q_1(K, L)) = \pi_F P_1(F) \leftrightarrow C \quad (7)$$

$$\pi_K(P_1(F) \times Q_1(K, L)) = \pi_K Q_1(K, L) \leftrightarrow \exists R. T \quad (8)$$

$$\pi_L(P_1(F) \times Q_1(K, L)) = \pi_L Q_1(K, L) \leftrightarrow \exists R^-. T \quad (9)$$

$$\pi_{K,L}(P_1(F) \times Q_1(K, L)) = \pi_{K,L} Q_1(K, L) \leftrightarrow R \quad (10)$$

$$\pi_{K,L}(P_1(F) \times Q_1(K, L)) = \pi_{K,L} Q_1(K, L) \leftrightarrow R \quad (11)$$

$$\pi_{F,L}(P_1(F) \times Q_1(K, L)) = P_1(F) \times \pi_L Q_1(K, L) \leftrightarrow \text{id}(C) \circ U \circ \text{id}(\exists R^-. T) \quad (12)$$

Добуток двох бінарних відношень необхідно спроекувати на один чи два атрибути, щоб отримати допустиме результуюче відношення. Дане відображення зводиться до варіанту добутку унарного та бінарного відношень:

- проекція на один атрибут зводиться до варіантів (8) та (9);
- проекція на два атрибути однакового відношення зводиться до (10);
- проекція на два атрибути різних відношень зводиться до (11) або (12).

Операція інверсного ділення

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

Операція номіналу

В дескриптивних логіках номінал – це конструктор концепту, який буде концепт із індивіду. Якщо d – ім'я індивіду, то $\{d\}$ є концепт. В RA^2 ця операція введена для підтримки конструктора номіналу.

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

Операція транзитивного замикання була включена в RA^2 для підтримки відповідного конструктору DL , тому:

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

В четвертому розділі запропоновано спосіб практичної перевірки методу відображення між дескриптивною логікою та бінарною реляційною моделлю даних на рівні RDF.

У підрозділі 4.1 наведено огляд літератури, що стосується використання RDF при відображенні реляційних баз даних в онтологію та онтології в реляційні бази даних. Цей огляд доповнює таксономію досліджень, що була створена у підрозділі 1.3 (рис. 1).

Підрозділ 4.2 присвячено постановці задачі практичної перевірки теоретичних результатів відображення $DL-RM^2$ та RM^2-DL . На рис. 3 наведено схему апробації методу відображення між DL та RM^2 .

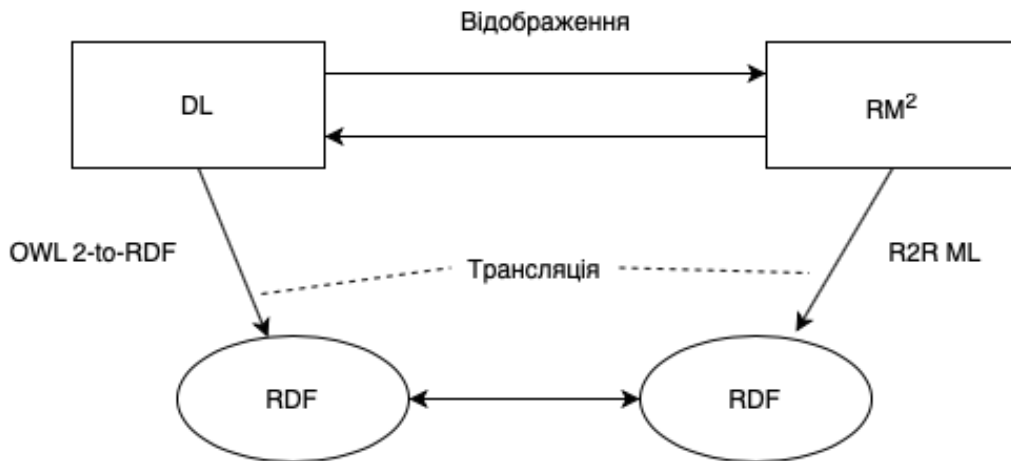


Рис. 3. Схема апробації методу відображень між DL та RM^2

Ідея апробації полягає в тому, щоб здійснити перевірку теоретичних результатів за допомогою методів та засобів мови RDF.

Відображення $DL-RM^2$ транслюються у RDF за допомогою мови перетворення R2R ML. Для виконання трансляції за результатами відображення створюється база даних, що ґрунтується на RM^2 . Відображення RM^2-DL транслюються у RDF за допомогою правил трансляції OWL 2-онтологій у RDF, які викладено у стандарті «OWL 2 Web Ontology Language Mapping to RDF Graphs». Для виконання трансляції за результатами відображення створюється онтологія предметної області на підставі тих даних, що існували у вхідній RDM. Перед здійсненням відображення RDM декомпозується на сукупність відношень арності не більше 2 за допомогою алгоритму, розглянутому у підрозділі 2.2.

Підрозділ 4.3 присвячено огляду засобу трансляції OWL 2-RDF. Визначено область обмеження способу перевірки, яка існує з огляду на структуру мови опису онтології. Відомо, що в основі OWL 2 лежить дескриптивна логіка SROIQ. Таким чином, область перевірки відображень за допомогою трансляції OWL 2-to-RDF обмежена лише тими операціями, які представлені у DL SROIQ.

Підрозділ 4.4 присвячено огляду мови R2R ML, його методам та засобам здійснення трансляції реляційних баз даних (RDB) у RDF-трійки. У R2R ML існує певний набір особливостей. Він дозволяє перетворювати структуру та обмеження цілісності RDB у RDF-трійки. Однак, будь-яку операцію можна відобразити лише у якості частини або цілковито SQL запиту. Кожен SQL запит представляється у якості логічної таблиці всередині карти трійок, що згенеровано для такої таблиці. Таким чином, не існує механізмів самостійного прямого перетворення кожної окремої операції реляційної алгебри всередині суто R2R ML.

Підрозділ 4.5 присвячено апробації алгоритму відображення $DL-RM^2$ у RDF. Практичну перевірку виконано на тестових даних (онтологія). Відображення здійснено у бінарну реляційну базу даних, яка створена з використанням концептуальної інформаційної моделі DL.

Обґрунтовано використання наступних правил перетворення.

Правило 1 (перетворення концептів). Кожен концепт (клас онтології), який утворюється на проміжному етапі створення складеного концепту (класу)

онтології) поійменовано та відображено у БД таким самим чином, як і атомарні концепти (класи онтології).

Правило 2 (перетворення конструкторів). Кожен конструктор концепту в онтології OWL 2 поійменовано та реалізовано у вигляді обмеження. Кожне обмеження відображено у БД таким самим чином, як і атомарні ролі (властивості онтології).

Результати апробації свідчать що алгоритм відображення утворює RDF-граф, який є еквівалентним RDF-графу, що утворюється за допомогою прямого перетворення онтології на RDF-трійки.

Підрозділ 4.6 присвячено апробації механізму відображення RM^2 -DL у RDF. Практичну перевірку виконано на тестових даних (база даних), яку було декомпозовано на сукупність унарних та бінарних відношень. Результат декомпозиції було відображено в онтологію, яку перетворено на RDF-граф. Результати апробації свідчать що алгоритм відображення утворює RDF-граф, який є еквівалентним RDF-графу, що утворюється за допомогою прямого перетворення БД на RDF-трійки.

Підрозділ 4.7 присвячено оцінкам складності алгоритму відображення між дескриптивною логікою та реляційною моделлю даних.

Оскільки дескриптивна логіка ALC є PSPACE-повною, проблема обчислення запитів в реляційній алгебрі також є PSPACE-повною, а трансляція дескриптивної логіки ALC в реляційну алгебру і навпаки представляє собою поліноміальну редукцію, то проблема функціонування системи, що підтримує відображення DL у реляційну алгебру та навпаки є PSPACE-повною. Що стосується включення транзитивних ролей, то питання складності функціонування системи, що підтримує відображення DL у реляційну алгебру та навпаки, залишається відкритим.

ВИСНОВКИ

Основним результатом дисертаційної роботи є розробка методу відображення між дескриптивною логікою та реляційною моделлю даних. В рамках проведених досліджень було отримано наступні наукові результати.

1) Побудовано таксономію досліджень, які стосуються відображень між онтологіями та реляційними базами даних; визначено вузькі місця методологій, з яких складається визначена таксономія; обґрунтовано важливість опису відображень між дескриптивною логікою та реляційною моделлю даних, а також вибір централізованої схеми інтеграції даних IC, керуючись результатами проведеного аналізу проблеми інтеграції даних в семантичному веб.

2) Уперше побудовано концептуальну інформаційну ER-модель дескриптивної логіки із залученням технології, яка забезпечує: підтримку гіпотези відкритого світу та не унікальності імен, повне представлення усіх інформаційних складових дескриптивної логіки, використання тільки унарних та бінарних сутностей, безпосереднє її відображення у бінарну реляційну структуру даних зі схемою бази даних у третій нормальній формі.

3) Визначено *спеціальний* варіант класичної RDM – бінарну реляційну модель даних (RM^2) яка складається з бінарної реляційної структури даних (RS^2), бінарної реляційної алгебри (RA^2), та обмежень цілісності. Оскільки, будь-яке n -арне відношення можна представити сукупністю бінарних, було запропоновано спосіб перетворення n -арної (класичної) реляційної структури в бінарну.

4) Уперше визначено бінарну реляційну алгебру (RA^2), у яку внесено принципово нові операції по відношенню до класичної реляційної алгебри (RA), для підтримки відображення конструкторів концептів та ролей DL. Також, у RA^2 внесено операції RA такі, що результат їх виконання утворює бінарне чи унарне відношення. Операції RA , що утворюють відношення арністю більше 2, були внесені в RA^2 у модифікованому вигляді, такому, що результат їх виконання теж утворює унарне або бінарне відношення.

5) Уперше розроблено метод двостороннього відображення між онтологічною та реляційною моделями даних, який: носить формальний характер, оскільки дескриптивна логіка (формальний апарат онтологій) розглядається у якості самостійної моделі даних; визначається як на структурному, так і на операційному рівні (включено конструктори концептів та ролей, а також операції реляційної алгебри); передбачає відображення термінологічних аксіом дескриптивної логіки.

6) Проведено програмну апробацію алгоритму відображення за допомогою мови RDF на прикладі тестових даних; визначено алгоритмічну оцінку складності.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Andon P.I. Reznichenko V.A., Chystiakova I.S. Mapping of description logic to the relational data model. *Cybernetics and Systems Analysis*. 2017. Vol. 6 no 53, pp. 160–175.
2. Чистякова І.С. Роль онтологій в інтеграції даних в семантичному вебе. *Інженерія програмного забезпечення*. 2014. № 1(17). С. 52–63.
3. Чистякова І.С. Інженерія онтологій. *Інженерія програмного забезпечення*. 2014. № 4(20). С. 53–68.
4. Резниченко В.А., Чистякова І.С. Отображення дескриптивної логіки АЛС в бінарну реляційну структуру даних. *Проблеми програмування*. 2015. № 4. С. 13–30.
5. Чистякова І.С. Інтеграція логік з операціями над ролями з реляційною моделлю даних. *Проблеми програмування*. 2016. № 4. С. 58–65.
6. Чистякова І.С. Інтеграція аксіоматики дескриптивних логік з реляційною моделлю даних. *Проблеми програмування*. 2017. № 1. С. 51–58.
7. Резниченко В.А., Чистякова І.С. Бінарна реляційна модель даних. *Проблеми програмування*. 2017. № 2. С. 96–105.
8. Chystiakova I.S. Implementation of mappings between the description logic and the binary relational data model on the RDF level. *Problems in programming* 2020. № 4. С. 41–54.
9. Chystiakova I.S. Mapping of the descriptive logic into RDF using binary relational data model. *Problems in programming* 2021. № 1. С. 56–83.
10. Чистякова І.С. Онтолого-орієнтована інтеграція даних в семантичному вебе. *Проблеми програмування*. 2014. № 2-3. Спец. випуск. Матеріали ІХ Міжнародної науково-практичної конференції з програмування «УкрПРОГ'2014». С. 188–196.
11. Chystiakova I.S. Methodologies for Ontology Creation. Proceedings of the sixth world congress «Aviation in the XXI-st century». (September 23-25, 2014, Kyiv, Ukraine). Київ, 2014. С. 1.13.5 – 1.13.11.
12. Резниченко В.А., Чистякова І.С. Інтеграція семейства розширених дескриптивних логік з реляційною моделлю даних. *Проблеми програмування*. 2016. № 2-3. Спец. випуск. Матеріали Х Міжнародної науково-практичної конференції з програмування «УкрПРОГ'2016». С. 38–47.
13. Чистякова І.С. Отображення реляційної алгебри в дескриптивну логіку. *Проблеми програмування*. 2018. № 2-3. Спец. випуск. Матеріали ХІ Міжнародної науково-практичної конференції з програмування «УкрПРОГ'2018». С. 214–225.
14. Chystiakova I. S. Implementation method of mappings between the descriptive logic and the binary relational data model by RDF. Матеріали ІХ Міжнародної науково-практичної конференції «Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні '2021» (Київ, 13-14 травня 2021). – Київ, 2021. – с.115–117.

АНОТАЦІЯ

Чистякова І.С. Метод відображення між дескриптивною логікою та реляційною моделлю даних. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 01.05.03 – «Математичне та програмне забезпечення обчислювальних машин і систем». – Київський національний університет імені Тараса Шевченка, Київ, 2021.

Дисертаційна робота присвячена дослідженню та розробці методу відображення між моделлю даних дескриптивною логікою та реляційною моделлю даних, а також їх апробації у прикладних задачах.

Основні результати роботи полягають у наступному: побудовано концептуальну інформаційну ER-модель дескриптивної логіки із залученням технології, яка забезпечує: підтримку гіпотези відкритого світу та не унікальності імен, повне представлення усіх інформаційних складових дескриптивної логіки, використання тільки унарних та бінарних сутностей; визначено спеціальний варіант класичної RDM – бінарну реляційну модель даних (RM^2) яка складається з бінарної реляційної структури даних (RS^2), бінарної реляційної алгебри (RA^2), та обмежень цілісності; запропоновано спосіб перетворення n-арної (класичної) реляційної структури в бінарну; визначено бінарну реляційну алгебру (RA^2), у яку внесено принципово нові операції по відношенню до класичної реляційної алгебри (RA), операції RA такі, що результат їх виконання утворює бінарне чи унарне відношення та операції RA , що утворюють відношення арністю більше 2, внесено у модифікованому вигляді, такому, що результат їх виконання теж утворює унарне або бінарне відношення; розроблено метод двостороннього відображення між онтологічною та реляційною моделями даних, який носить формальний характер, визначається як на структурному, так і на операційному рівні і передбачає відображення термінологічних аксіом дескриптивної логіки.

Ключові слова: відображення, онтологія, реляційна модель даних, бінарна реляційна модель даних, бінарна реляційна алгебра, дескриптивна логіка, ALC, RDF, OWL 2, R2R ML.

АННОТАЦИЯ

Чистякова И.С. Метод отображения между дескриптивной логикой и реляционной моделью данных. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.03 – «Математическое и программное обеспечение вычислительных машин и систем». – Киевский национальный университет имени Тараса Шевченко, Киев, 2021.

Диссертационная работа посвящена исследованию и разработке метода отображения между моделью данных дескриптивной логикой и реляционной моделью данных, а также их апробации в прикладных задачах.

Основные результаты работы заключаются в следующем: построено концептуальную информационную ER-модель дескриптивной логики с привлечением технологии, которая обеспечивает: поддержку гипотезы открытого мира и не уникальности имен, полное представление всех

информационных составляющих дескриптивной логики, использование только унарных и бинарных сущностей; определен специальный вариант классической RDM – бинарную реляционную модель данных (RM^2), которая состоит из бинарной реляционной структуры данных (RS^2), бинарной реляционной алгебры (RA^2), и ограничений целостности; предложен способ преобразования n-арной (классической) реляционной структуры в бинарную; определено бинарную реляционную алгебру (RA^2), в которую внесены принципиально новые операции по отношению к классической реляционной алгебре (RA), операции RA такие, что результат их выполнения образует бинарное или унарное отношение, а также операции RA, образующие отношение арностью более 2, внесены в модифицированном виде, таком, что результат их выполнения тоже образует унарное или бинарное отношение; разработан метод двустороннего отображения между онтологической и реляционной моделями данных, который носит формальный характер, определяется как на структурном, так и на операционном уровне и предусматривает отображение терминологических аксиом дескриптивной логики.

Ключевые слова: отображения, онтология, реляционная модель данных, бинарная реляционная модель данных, бинарная реляционная алгебра, дескриптивная логика, ALC, RDF, OWL 2, R2R ML.

ABSTRACT

Chystyakova I.S. Mapping method between the description logic and the relational data model. Manuscript.

The thesis for a candidate of technology science degree in specialty 01.05.03 – «Mathematical and Software of Computers and Systems». – Taras Shevchenko National University of Kyiv, Kyiv, 2021.

The thesis is dedicated to the research and development of a mapping method between the description logic data model and the relational data model as well as their approbation in the applied tasks.

The main results of the work are as follows: a conceptual information ER-model of description logic was built using technology that provides: support for the open world assumption and non-uniqueness of names, a complete representation of all information components of description logic, the use of only unary and binary entities; the binary relational data model (RM^2) is created, which is a special version of the classical RDM; the binary relational data structure (RS^2) is described; a way to transform the n-ary (classical) relational structure into the binary is proposed; the binary relational algebra (RA^2) is proposed; the RA^2 operations, which result is an n-ary relation, are contributed into the RA^2 in a modified form and have an unary or binary relation in result; a method of two-way mapping between ontological and relational data models has been developed, which is of a formal nature, is defined both at the structural and operational levels and provides for the mapping of the terminological axioms of description logic.

Keywords: mappings, ontology, relational data model, binary relational data model, relational algebra, description logic, ALC, RDF, OWL 2, R2R ML.