Мохаммед Карам Джасім, тимчасово не працює: &laquo;Роз&shy;ширення табличних алгебр множинним успадкуванням&raquo; (01.05.03 - математичне та програмне забезпечення об&shy;числювальних машин і систем). Спецрада Д 26.001.09 у Київському національному університеті імені Тараса Шевченка

Київський національний університет імені Тараса Шевченка

Кваліфікаційна наукова

праця на правах рукопису

Мохаммед Карам Джасім

УДК 004.655

ДИСЕРТАЦІЯ

Розширення табличних алгебр множинним успадкуванням.

01.05.03 – Математичне та програмне забезпечення обчислювальних машин і

систем

12 – Інформаційні технології

Подається на здобуття наукового ступеня кандидата фізико-математичних наук

Дисертація містить результати власних досліджень. Використання ідей,

результатів і текстів інших авторів мають посилання на відповідне джерело

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис, ініціали та прізвище здобувача)

Науковий керівник Буй Дмитро Борисович,

доктор фіз.-мат. наук, професор

Київ – 2017

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ....................... 17

ВСТУП.............................................................................................................. 20

ЗМІСТОВНА СЕМАНТИКА МОВИ SQL ................................................ 27

1.1 Огляд мови SQL.........................................................................27

1.2 Опис бази даних інвойсів..........................................................30

1.3 Базовий синтаксис оператору запитів SELECT......................37

1.4. Сортування рядків ....................................................................46

1.5 Операції зовнішнього та внутрішнього з’єднання .................48

РОЗДІЛ 2. УТОЧНЕНА ТАБЛИЧНА АЛГЕБРА СЕМАНТИЧНИХ

ФУНКЦІЙ ЗАПИТІВ ЯДРА МОВИ SQL.................................................. 53

2.1. Математичні основи сучасних реляційних СУБД ................53

2.2. Визначення основних структур даних та операцій на них...55

2.3 Теоретико-множинні операції на таблицях ............................58

2.4 Операції з’єднання таблиць ......................................................61

2.5 Визначення операторів на таблицях ........................................62

2.6. Відношення конфінальності, передпорядки та порядки ......67

2.7 Семантика фрази ORDER BY запитів SQL-подібних мов ....79

РОЗДІЛ 3. МАТЕМАТИЧНІ ОСНОВИ АЛГОРИТМІВ

ЛІНЕАРИЗАЦІЇ.............................................................................................. 82

3.1 Основні властивості операції лінеаризації..............................82

3.2 Допоміжні операції інверсії і добутку бінарних відносин ....83

3.3. Допоміжні результати: рефлексивність, антисиметричність і

транзитивність..................................................................................85

3.4. Рефлексивно-транзитивне замикання бінарного

відношення .......................................................................................88

3.5 Рефлексивно-транзитивне замикання бінарного відношення:

характеристичні денотатівні властивості......................................91

16

3.5.1 Рефлексивно-транзитивне замикання як найменше

рефлексивне і транзитивне відношення, що містить вихідне

відношення...............................................................................................91

3.5.2 Рефлексивно-транзитивне замикання як подалгебра, яка

породжена діагоналлю в спеціальній алгебри всіх пар.......................94

3.6 Рефлексивно-транзитивне замикання бінарного відношення

як найменше рішення характеристичного рівняння ....................97

3.7 Основні результати і висновки...............................................100

РОЗДІЛ 4. МНОЖИННЕ УСПАДКУВАННЯ ТАБЛИЦЬ З

ВИКОРИСТАННЯМ АЛГОРИТМІВ ЛІНЕАРИЗАЦІЇ....................... 102

4.1 Основні визначення та поняття успадкування таблиць.......102

4.2 Одиночне успадкування таблиць ...........................................107

4.3 Множинне успадкування таблиць та класів .........................113

4.4 Приклади множинного успадкування таблиць.....................122

ВИСНОВКИ .................................................................................................. 130

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ................................................. 131

Додаток 1 База даних інвойсів................................................................... 146

Додаток 2. Список публікацій здобувача ................................................ 156

Додаток 3. Відомості про апробацію......................................................... 159

17

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ



порожня множина



def

за означенням



def

дорівнює за означенням

O

доповнення множини

O



відношення конфінальності на множинах

P Q x x P y y Q x y

def

  (    (  &  ))



~

узагальнена рівність

U|X

обмеження бінарного відношення

U

за множиною

X

U[ X ]

повний образ множини

X

відносно бінарного відношення

U

[a, x)

початковий відрізок ц. у. м.

~

стрілка для запису часткових операцій

U V

композиція бінарних відношень

rang

область значень операції



i

n

проекція множини кортежів довжини

n

за

i -тою

компонентою

1,n

множина натуральних чисел

{1,...,n}

( )

~

DD

сім’я часткових функцій з множини

D

у множину

D



бінарне відношення сумісності функцій

dom f

область означеності функції

f

true

істинне булеве значення

false

хибне булеве значення

P( X )

булеан множини

X

18

2

X

множина всіх скінченних підмножин множини

X

U

1

бінарне відношення, обернене (інверсне) відношенню

U

R

унарна композиція рекурсії для випадку одного параметра

(D ... D D D) n

m

n

1

1

   



підклас функцій класу

(D ... D D D) 1  n  

, монотонних

за

n  1-им аргументом;

D – ч. в. м.

X

Y

множина всіх скінченних функціональних відношень на

парі множин

Y, X





\*

транзитивне замикання відношення





X|U

приведення бінарного відношення

U

за множиною

X

|A|

носій алгебри

[]

замикання множини



операціями сім’ї





порожня іменна множина (розд. 4); рядок порожньої

схеми (розд. 5)



рефлексивно-транзитивне замикання відношення



(d ,d ) 1 2

іменна множина зі стандартними іменами

{ 1,d2 , 2,d2 }



об’єднання сумісних іменних множин



об’єднання

s-множин

\

різниця

s-множин

n

сигнатура програмних алгебр іменних функцій

 v

параметрична функція іменування,

v – ім’я

v 

параметрична функція розіменування за іменем

v



іменний предикат рівності іменних даних

 R

,  R

, \R

відповідно об‘єднання, перетин, різниця таблиць схеми R

 з‘єднання, еквіз‘єднання, внутрішнє природне з‘єднання

 R

R

2

1

ділення таблиць схеми

R1

на таблиці схеми

R2

19

t

порожня таблиця

t

таблиця, що містить єдиний рядок



порожньої схеми;

t

def



 {}

~

відношення односхемності (таблиць),

t t R t T R t T R

def

1 2 1 2 ~  (  ( )&  ( )) ~

S

відношення односхемності рядків

t t t

, ,\

відповідно операції об‘єднання, перетину, різниці

односхемних таблиць

Cj

операція декартова з‘єднання (cross join, Cartesian join)



A1 An

,...,

операція внутрішнього з’єднанням за атрибутами

A1 An

,..., , n  1

(inner join using

A1 An

,...,

)



p

операція внутрішнього з’єднанням за предикатом

p

(inner

join on

p

)

S

тризначний предикат рівності мови SQL

m

операція декартова з‘єднання мультимножин

OrderbyA



параметрична функція побудови у-таблиць

СУБД система управління базою даних

20

ВСТУП

Перші реляційні СУБД були розроблені ще в 70-ті роки минулого століття.

На початку 90-х років, до реляційних баз даних були застосовані ідеї об’єктноорієнтованого підходу, що привело до появи об’єктно-реляційних систем

управління базами даних (ОРБД), які поєднують функціональність реляційних

СУБД з концепціями та ідеями взятими з об’єктно-орієнтованого

програмування. В стандартах SQL:1999 - SQL:2016 в реляційну модель даних

були додані численні об’єктно-орієнтовані риси. Це значно змінило модель

даних, що лежить в основі останніх версій СУБД, базованих на цих стандартах.

На сьогодні, незважаючи на існування інших моделей даних[80, 94, 121, 30],

реляційні та об’єктно-реляційні бази даних залишаються основою

інформаційних систем підприємств та організацій.

Семантиці SQL присвячені численні теоретичні роботи, починаючи,

наприклад, з роботи в якій була побудована формальна модель запитів, названа

E3VPC [36], та закінчуючи публікацією в 2017 році денотаційної семантики SQL,

названої HoTTSQL [46] Потужність реляційних СУБД значною мірою

пояснюється тим, що вони базуються на строгих математичних основах,

викладених ще Кодом [12, 13]. В той же час, всі ці роботи не приділяють уваги

об’єктно-орієнтованим властивостям СУБД, зосередившись, на структурах

даних та операціях над ними.

На кафедрі теорії та технології програмування Київського національного

університету імені Тараса Шевченка розвивається побудова семантики SQL на

основі спеціальної програмної алгебри, названої табличною алгеброю

семантичних функцій [63, 26, 60, 115, 50, 49, 53, 54, 55, 56, 97, 57, 94, 96, 99, 100,

93, 95], яка має ряд переваг над іншими підходами. А саме, в табличній алгебрі

виділяється два типи функцій – функції на даних, зокрема таблицях, та функції

на функціях, які називаються операторами, як це прийнято в інших мовах

програмування. Семантика складних виразів SQL в цій алгебрі природнім чином

будується з функцій на даних і операторів.

21

Слід відмітити, що такий підхід до визначення семантики мов

програмування був запропонований в роботах по системам алгоритмічних

алгебр (САА) В.М. Глушкова, Г.Е. Цейтліна, Е.Л. Ющенко[65], а згодом в

роботах по композиційній семантиці В.Н. Редька, Д.Б. Буя та М.С. Нікітченка

[102, 106, 107, 108, 109, 110]. Ці праці стали теоретичним підґрунтям Київської

школи програмування (http://www.computer-museum.ru/galglory/27.htm).

Потреба у розвитку теорії до рівня сучасних ОРБД, зокрема семантики

запитів мови SQL, та семантики об’єктних розширень реляційної моделі даних,

і визначає актуальність даної дисертаційної роботи в який розширюється

таблична модель даних за рахунок множинного успадкування таблиць, та

удосконалена таблична алгебра семантичних функцій SQL, зокрема

перевизначені оператори внутрішнього та зовнішнього з’єднання.

Зв’язок роботи з науковими програмами, планами, темами.

Дисертаційна робота є складовою частиною наукових робіт, проведених на

кафедрі теорії та технології програмування факультету комп’ютерних наук та

кібернетики Київського національного університету імені Тараса Шевченка при

виконанні фундаментальної теми "Формальні специфікації та методи розробки

надійних програмних систем" (№ 0111U007052, 2011-2015 рр.) і теми «Теорія і

методи розробки інтелектуальних інформаційних технологій та систем»

(№16КФ015-02).

Мета і задачі дисертаційного дослідження. Метою дисертаційної роботи

є аналіз алгебри семантичних функцій ядра мови SQL, яка називається

табличною алгеброю, на відповідність реалізаціям SQL в поширених СУБД,

зокрема операцій з’єднання та теоретико-множинних операцій; знаходження

розбіжностей, тобто випадків, коли результати обчислень семантичних функцій

відрізняються від результатів, які повертають відповідні запити реалізовані в

мові SQL; модифікація алгебри семантичних функцій для уникнення таких

розбіжностей; розширення реляційної моделі даних конструкцією одиночного і

множинного успадкування таблиць.

22

З огляду на мету в роботі ставляться такі задачі:

провести перевірку табличної алгебри ядра мови SQL. Перевірку провести

на версії SQL, втіленої в об’єктно-реляційної СУБД PostgreSQL, яка є однією з

найбільш точних та повних реалізацій стандарту SQL. При цьому особу увагу

приділити обчисленню запитів на таблицях, які мають співпадаючи імена

стовбців, або які є пустими

перевизначити семантику операторів внутрішнього та зовнішнього

з’єднання реляцій таким чином, щоб результати з’єднання для реляцій, імена

атрибутів яких частково або повністю співпадають, повертали такі самі реляції

як і в мові SQL, по можливості, усунути також інші розбіжності

перевірити семантичні функції, які задають семантику фрази ORDER BY.

Дослідити властивості семантичних функцій, які відповідають за впорядкування

рядків таблиць

дослідити алгоритми автоматичного вирішення конфліктів імен атрибутів

при множинному успадкуванні, та застосувати ці алгоритми до успадкування

таблиць

дослідити властивості методу лінеаризації, який використовується в

алгоритмах множинного наслідування

Предметом дослідження є реляційна модель даних.

Об’єктом дослідження є реляційні бази даних та об’єктно-реляційні бази

даних.

Наукова новизна одержаних результатів. Проведене порівняння

результатів запитів ядра мови SQL та результатів обчислень відповідних їм

семантичних функцій табличної алгебри. Виявлені розбіжності в результатах

виконання запитів, які містять операції зовнішнього та внутрішнього з’єднання

та відповідним їм семантичним функціям на таблицях, які мають спільні імена

атрибутів. Також виявлені розбіжності при виконанні теоретико-множинних

23

операцій на таблицях, пов’язані з тим, що в SQL операції об’єднання, перетину

та різниці таблиць враховують порядок стовбців заданий в запитах, а в

семантичній алгебрі враховуються імена атрибутів стовбців, а порядок не є

важливим.

Запропоновані нові визначення для операцій внутрішнього та зовнішнього

з’єднання та теоретико множинних операцій, визначена нова форма оператору

суперпозиції, більш зручна для задання семантики складних запитів. В

результаті, поновлена алгебра семантичних функцій більш точно описує

семантику ядра мови SQL та є більш зручною в використанні.

Проведена перевірка семантики фрази ORDER BY мови SQL. Виявлено,

що семантичні функції задають частковий порядок на кортежах, в той час як

фраза SQL ORDER BY завжди повертає лінійний порядок на рядках таблиці.

Сформульована та доведена теорема, яка визначає умови, при яких семантичні

функції задають точну семантику ORDER BY.

Введено поняття ієрархії успадкування, яке описує як одиночне так і

множинне успадкування таблиць. Сформульовані та доведені теореми про

характеристичні ознаки одиночного успадкування. Надані формальні

математичні визначення для таких властивостей методу лінеаризації ієрархії

таблиць, як монотонність та збереження локального порядку успадкування.

Що стосується практичної частини, то розроблені алгоритми перевірки, чи

є ієрархія успадкування одиночним успадкуванням чи множинним, та чи є

ієрархія одиночного успадкування простою.

Надане формальне математичне визначення алгоритму MRO C3 для

автоматичного вирішення конфлікту співпадаючих імен атрибутів при

множинному успадкуванні таблиць. Доведені теореми про монотонність

алгоритму та його завершуваність.

24

Теоретичне і практичне значення одержаних результатів. Дисертація

має теоретико-прикладну спрямованість. Результати роботи можуть

використовуватись для задання семантики запитів мови SQL, для оптимізації

запитів, досліджені властивостей мови SQL, доведенні коректності запитів, для

еквівалентного перетворенні запитів та доведення еквівалентності запитів,

проектуванні баз даних.

Результати роботи були впроваджені у навчальний процес за

спеціальністю "Інформатика" на факультеті кібернетики КНУ (нормативні курси

"Прикладна логіка", "Композиційна семантика SQL-подібних мов"; спеціальний

курс "Вступ до реляційних баз даних"), а також у Кіровоградському державному

педагогічному університеті імені Володимира Винниченка (нормативний курс

"Табличні алгебри та SQL подібні-мови").

Особистий внесок здобувача. Всі результати, які складають суть

дисертаційної роботи, отримані здобувачем самостійно. З праць, виконаних зі

співавторами, на захист виносяться лише результати, отримані особисто

здобувачем. У спільно виконаних роботах науковому керівнику Д.Б. Бую

належить загальна постановка проблеми, обговорення та інтерпретація

результатів.

Апробація результатів дослідження. Основні положення та висновки

дисертаційного дослідження обговорювалися на наукових семінарах кафедри

теорії та технології програмування КНУ, республіканському семінарі

"Програмологія та її застосування".

Результати дисертаційного дослідження оприлюднені у доповідях і

повідомленнях на Міжнародних та Всеукраїнських наукових конференціях,

семінарах:

25

XI Міжнародна науково-практична конференція: ТЕОРЕТИЧНІ ТА

ПРИКЛАДНІ АСПЕКТИ ПОБУДОВИ ПРОГРАМНИХ СИСТЕМ. – Київ, 15-17

грудня 2014 р.

І Международный научно-практический форум «Наука и бизнес»: –

Черновцы, 2015.

13th International Scientific Conference on Informatics. – Poprad, Slovakia, 18-

20 Nov., 2015.

XII Міжнародна науково-практична конференція ТЕОРЕТИЧНІ ТА

ПРИКЛАДНІ АСПЕКТИ ПОБУДОВИ ПРОГРАМНИХ СИСТЕМ. – Київ, 23-26

грудня 2015 року.

XIII Міжнародна науково-практична конференція ТЕОРЕТИЧНІ ТА

ПРИКЛАДНІ АСПЕКТИ ПОБУДОВИ ПРОГРАМНИХ СИСТЕМ. – Київ, 5-9

грудня 2016 року.

XII International Conference on Perspective Technologies and Methods in

MEMS Design (MEMSTECH). – 2016.

Міжнародний науково-практичний семінар – "Комбінаторні конфігурації

та їх застосування" – 2016.

Четвёртая конференция «КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ В

НАУКОЕМКИХ ТЕХНОЛОГИЯХ» (КМНТ-2016). – Харьков, 26-31 мая 2016 г

Публікації. Результати дисертації опубліковані у 7 статтях [81, 82, 83, 84,

89, 90, 32] в наукових журналах і збірниках наукових праць, 8 тезах конференцій

[91, 87, 33, 88, 86, 35, 85, 34]; з них 5 статей опубліковані у фахових виданнях,

затверджених ВАК України, одна стаття опублікована в міжнародному журналі.

Структура та обсяг дисертації. Дисертаційна робота з анотації, переліку

умовних позначень, вступу, чотирьох розділів, висновків, списку використаних

26

джерел. Загальний обсяг дисертації становить 160 с, основний текст 110 с.,

список використаних джерел 121 найменування на 15 сторінках.

ВИСНОВКИ

Головнимрезультатомдисертаціїєрозширеннятабличноїалгебри

семантичнихфункціймовиодиночнимтамножиннимуспадкуванням

таблицьзавтоматичнимрозв’язаннямконфліктівіменатрибутівзадопомогою

алгоритмулінеаризаціївідомогопідназвоюЦерозширеннярозв’язує

важливузадачупобудовиматематичноїмоделіоб’єктнореляційнихбазданих

щомаєвеликетеоретичнетапрактичнезначеннядлярозробкиперспективних

комп’ютернихінформаційнихсистем

Основнірезультативключаютьвсебе

Длябільшточногоописусемантикиядрамовитабільшої

зручностіувикористаннізамістьстарихвизначеньвтабличнійалгебрі

семантичнихфункційзапропонованіновівизначенняоперацій

внутрішньоготазовнішньогоз’єднанняатакожтеоретикомножинних

операцій

Доведенатеоремаякавизначаєумовиприякихсемантичніфункції

задаютьточнусемантикуоператораЦепоказалощоповна

семантикаможебутизаданалишенедетермінованими

функціями

Доведенітеоремипрохарактеристичніознакиодиночного

успадкуванняНаїхбазірозробленіалгоритмиперевіркичиєієрархія

успадкуванняодиночнимчимножиннимуспадкуваннямтачиє

ієрархіяодиночногоуспадкуванняпростою

Формальновизначенітаківажливівластивостіметодулінеаризації

ієрархіїтаблицьякмонотонністьтазбереженнялокальногопорядку

успадкуваннящодаєможливістьоцінититойчиіншийметод

лінеаризаціїтавизначитинаскількивінєбезпечним



Побудованоалгоритмдляавтоматичноговирішення

конфліктуспівпадаючихіменатрибутівпримножинномууспадкуванні

таблицьЦедозволилоперевірятивластивостіалгоритмунастрогому

математичномурівніЗокремадоведенотеоремипротаківластивості

алгоритмуякмонотонністьтайогозавершуваністьАлгоритм

розроблювавсянапсевдокодіщозробилойогонезалежнимвідмови

програмуванняДодаткововінбувреалізованийнамовіпрограмування

Проведенетестуванняпрограмноїреалізаціїтакожпідтвердило

коректністьдоведенихтеорем