

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

ГОРБАНЬ ГЛІБ ВАЛЕНТИНОВИЧ



УДК 004.652.5

**МЕТОДИ ТА ОБ'ЄКТНО-ОРІЄНТОВАНА
ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ БАГАТОМІРНИХ ДАНИХ**

Спеціальність 05.13.06 – інформаційні технології

Автореферат дисертації
на здобуття наукового ступеня
кандидата технічних наук

Миколаїв – 2016

Дисертацією є рукопис

Робота виконана на кафедрі інтелектуальних інформаційних систем у Чорноморському державному університеті імені Петра Могили Міністерства освіти і науки України

Науковий керівник: доктор технічних наук, професор
Фісун Микола Тихонович,
Чорноморський державний університет
ім. Петра Могили,
завідувач кафедри інтелектуальних
інформаційних систем

Офіційні опоненти: доктор технічних наук, професор
Філатов Валентин Олександрович,
Харківський національний університет
радіоелектроніки,
завідувач кафедри штучного інтелекту;

доктор технічних наук, професор
Шаховська Наталія Богданівна,
Національний університет
«Львівська політехніка»,
професор кафедри інформаційних систем та мереж

Захист відбудеться «12» квітня 2016 року о 14:00 на засіданні спеціалізованої вченої ради Д 38.053.05 у Чорноморському державному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68 Десантників, 10.

З дисертацією можна ознайомитися в бібліотеці Чорноморського державного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68 Десантників, 10.

Автореферат розісланий «11» березня 2016 р.

Вчений секретар
спеціалізованої вченої ради



М. П. Мусієнко

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

У сучасних інформаційно-аналітичних системах та системах підтримки прийняття рішень часто застосовуються технології OnLine Analytical Processing (OLAP) і Data Mining (DM). Вони дозволяють тією чи іншою мірою ефективно проводити аналіз даних. Технологія OLAP дозволяє користувачеві проводити різні аналітичні операції, такі як консолідація, деталізація, зріз даних, обертання куба та інші. При цьому куби OLAP створюються в сучасних системах керування базами даних (СКБД) на основі реляційних таблиць баз даних (БД) і не обов'язково на фізичному рівні насправді представляють куб. Технологія Data Mining являє собою дослідження в накопичених даних деяких прихованих знань, які до цього не були відомі і можуть виявитися достатньо корисними в процесі аналізу даних, при цьому дані беруть із заздалегідь створених реляційних СКБД. Таким чином, у сучасних реляційних СКБД технології OLAP та Data Mining рухаються у різних напрямках, оскільки в OLAP увага приділяється виключно забезпеченню доступу до багатомірних даних, а методи Data Mining у більшості випадків працюють з реляційними таблицями, у яких, як правило, розміщені часові ряди. Інтеграція цих двох технологій підвищить функціональність інформаційних систем, що орієнтовані на аналіз даних та підтримку прийняття рішень.

Вагомий внесок у дослідження інформаційних технологій баз даних, OLAP та Data Mining здійснено вітчизняними вченими, такими як П. І. Бідюк, О. А. Блажко, В. І. Гайдаржи, Ю. П. Зайченко, І. О. Кравець, О. Б. Кунгурцев, Є. В. Малахов, В. В. Пасічник, В. О. Філатов, М. Т. Фісун, Н. Б. Шаховська, та зарубіжними вченими, такими як А. А. Барсегян, Ю. О. Кудрявцев, С. Д. Кузнєцов, В. І. Орешков, М. Б. Паклін, А. О. Сахаров, С. Ю. Семченков, A. Abdulghani, R. Agraval, R. Ben Messaoud, O. Boussaid, G. Dong, S. Goil, J. Han, T. Imielinski, L. Khachiyani, A. Laurent, R. Loudcher, K. Parsaye, H. Pinto, S. Sarawgi, H. Zhu та ін.

Актуальність теми. Інформаційні технології баз даних, OLAP та Data Mining розвиваються у теперішній час достатньо потужно, але в інтеграційному плані існують досить окремо одна від одної. Аналіз джерел інформації дозволяє зробити висновок, що моделі та наведені вище інформаційні технології ще далекі від інтеграції в рамках єдиної інтелектуальної інформаційної системи, хоча окремо самі по собі технології OLAP та Data Mining реалізовані в багатьох реляційних СКБД (Microsoft SQL Server, Oracle, DB2, Informix та ін.). Їх можна розглядати як два взаємодоповнюючих напрями. З одного боку, OLAP-технологія має потужну здатність організувати структури даних, адаптовані до аналізу, але дана технологія обмежена простою навігацією, агрегуванням та деталізацією даних, що послаблює її здатність до глибокого аналізу. З іншого боку, технологія Data Mining має обмежені можливості для організації даних, але дана технологія відома своєю описовою та прогностичною силою, яка може виявити знання з простих і складних даних. Аналіз даних ставить нові завдання в технології баз даних. Їх комбінація приведе до другого покоління систем баз

даних, що дозволить створювати й управляти базами знань так само, як у класичних бізнес-застосуваннях.

У багатьох проектах з аналізу даних аналітики стикаються з великою кількістю варіантів у застосуванні методів інтелектуального аналізу даних. Вони включають в себе різні способи побудови моделей аналізу даних на по-різному сегментованих, агрегованих та трансформованих наборах даних. І доцільно було б об'єднати ці три технології у єдиній інформаційній системі. Така система підвищила б рівень інтелектуальності шляхом інтеграції вищезазначених моделей інформаційних технологій. Аналіз джерел інформації про дослідження щодо інтеграції OLAP і Data Mining, що реалізовані в реляційних СКБД, дозволяє дійти висновку, що вирішення цієї проблеми ще далеке від завершення. Тому тема дослідження, що спрямована на аналіз та інтеграцію названих моделей і технологій, представляє великий інтерес. Це дозволить підвищити коло задач систем підтримки прийняття рішень, що створюються у складі інтелектуальних інформаційних систем. І ще більш недослідженими такі питання є для постреляційних СКБД, що підтверджується досить малою кількістю публікацій стосовно цієї проблеми. Однією з таких постреляційних є об'єктна модель даних, яка знайшла своє відображення в об'єктних системах керування базами даних (ОСКБД). На ринку програмних продуктів уже з'явилися окремі ОСКБД, але технології OLAP і Data Mining у них не досягли, в силу більш пізньої появи в галузі ІТ, того рівня, що у реляційних СКБД. Тому тема дослідження, що стосується як подальшого розвитку моделей та інформаційних технологій OLAP і Data Mining в середовищі об'єктних СКБД, так і їх інтеграції, є *актуальною*.

Зв'язок роботи з науковими програмами, планами, темами. Дисертація виконана у відповідності до завдання НДР «Інтеграція інформаційних технологій баз даних, баз знань та Data Mining» (держ. рег. №0112U007749), у якій автор був відповідальним виконавцем, та «Інтелектуалізація інформаційних систем та СППР за рахунок впровадження методів ситуаційного моделювання, сценарного аналізу, Data Mining та OLAP технологій» (держ. рег. № 0114U004402), в якій автор брав участь як виконавець.

Мета та задачі дослідження. *Метою* роботи є підвищення рівня інтелектуальності інформаційних систем шляхом створення інструментальної програмної системи автоматизованого проектування об'єктних багатомірних баз даних, методів формування асоціативних правил за шаблонами та їх реалізації в складі системи, що дозволяє з набагато меншими витратами проектувати інформаційно-аналітичні системи в об'єктному програмному середовищі.

Для досягнення сформульованої мети в роботі треба вирішити наступні *задачі*:

- дослідити об'єктну модель даних і об'єктні бази даних з точки зору побудови системи OLAP та апробувати різноманітні алгоритми створення OLAP-кубів у середовищі ОСКБД;

- дослідити та розробити моделі, методи та алгоритми пошуку асоціативних правил в OLAP-кубах, вивести формули розрахунку характеристик їх значущості;
- спроектувати репозиторій метаданих для інтеграції технологій об'єктних баз даних, OLAP-кубів, асоціативних правил та перетворення результатів пошуку асоціативних правил у базу знань;
- дослідити та апробувати різноманітні об'єктні СКБД й обрати базове програмне середовище для створення інтелектуальної інформаційної системи;
- впровадити інтелектуальну інформаційну систему для створення новітніх систем накопичування, переробки, збереження та аналізу інформації.

Об'єкт дослідження: інтеграція моделей, методів та інформаційних технологій аналізу даних в інтелектуальних інформаційних системах та системах підтримки прийняття рішень.

Предмет дослідження: моделі, методи та інформаційні технології інтелектуальної обробки даних у середовищі об'єктних СКБД.

Методи дослідження. Наведені в дисертаційній роботі результати базуються на використанні теорії множин на етапі розробки моделі багатомірного куба, комбінаторики на етапі розробки методу обчислення агрегованих даних у багатомірному кубі, генерації шаблонів міжвимірових асоціативних правил у багатомірних даних та пошуку частих предметних наборів між різними вимірами у багатомірних кубах; методу Apriori на етапі генерації міжвимірових асоціативних правил у багатомірних даних, виходячи зі згенерованих їх шаблонів та отриманих частих предметних наборів; об'єктно-орієнтованої методології створення комп'ютерних систем на етапі проектування репозиторію метаданих для інформаційної системи, що об'єднує інформаційні технології об'єктних баз даних, OLAP та Data Mining у вигляді пошуку міжвимірових асоціативних правил у багатомірних кубах.

Наукова новизна отриманих результатів:

- уперше в рамках інформаційної системи запропоновано та досліджено модель інтеграції інформаційних технологій об'єктних баз даних, OLAP та Data Mining, яка дозволяє проводити інтелектуальний аналіз багатомірних даних на відміну тільки від одновимірного аналізу в реляційних системах;
- удосконалено модель представлення асоціативних правил, яка представляє їх у вигляді об'єктів класу, а також будує їх шаблони, до яких потім відносить відповідні правила;
- отримали подальший розвиток методи пошуку асоціативних правил у багатомірних даних, які, на відміну від існуючих, дозволяють знайти закономірності між фактами у розрізі різних вимірів куба;
- отримала подальший розвиток інформаційна технологія побудови структури метаданих для об'єктних СКБД, що полягає у її розширенні класами підсистем OLAP та Data Mining, котра дозволяє здійснювати

проектування довільної об'єктної бази даних на фізичному рівні та її безпосередній оперативний та інтелектуальний аналіз;

- удосконалено методи побудови OLAP-кубів, а саме – розроблено комбінаторний метод обчислення OLAP-куба, який, у порівнянні з іншими, дозволив проводити обчислення, що полягає в одноетапній процедурі обчислення агрегованих значень OLAP-куба за різними вимірами.

Практичне значення наукових результатів:

- розроблено метод побудови шаблонів для автоматизованого формулювання задач пошуку асоціативних правил;
- розроблено метакласи підсистем OLAP та Data Mining у складі об'єктної СКБД Caché;
- розроблено програмну інструментальну систему проектування об'єктної бази даних довільної предметної сфери як додаткову сервісну функцію (службу) об'єктної СКБД Caché;
- розроблені моделі, методи, алгоритми, інструментальні засоби та інформаційні технології використані у виробничій діяльності СП «ГРАСКО» (м. Миколаїв) (акт апробації від 10.12.2015);
- результати дисертаційних досліджень використовуються у навчальному процесі Чорноморського державного університету імені Петра Могили при викладанні дисципліни «Об'єктні бази даних» для студентів четвертого курсу напряму підготовки 6.050101 – Комп'ютерні науки (акт впровадження від 18.09.2015).

Особистий внесок здобувача. Наукові положення, теоретичні розробки та практичні результати у представленій дисертаційній роботі отримані автором особисто.

У наукових роботах, опублікованих у співавторстві, здобувачу належить: у [1] реалізовано алгоритм MultiValue для створення обчислення OLAP-кубів для постріляційної СКБД PostgreSQL; у [2] досліджена постріляційна СКБД Caché з точки зору побудови багатомірних структур даних; у [3, 10, 13] побудований багатомірний куб для оперативного аналізу даних мережевого трафіку мультисервісної корпоративної мережі; у [4] сформульовані, запрограмовані та протестовані результати реалізації функцій Data Mining у програмному середовищі СКБД MS SQL Server 2008; у [5, 12, 17] запропонована модель метаданих для побудови системи OLAP у середовищі об'єктних баз даних; у [6, 19] запропонована реалізація алгоритму DWARF для створення та обчислення OLAP-кубів для об'єктних баз даних; у [8, 9] запропонований підхід до генерації міжвимірових асоціативних правил у багатомірних даних шляхом знаходження всіх можливих їх шаблонів; у [9] запропонований метод знаходження частих предметних наборів у багатомірних даних шляхом їх представлення у вигляді багаторівневого списку.

Апробація результатів дисертації. Основні результати роботи були представлені на 18 міжнародних, 6 всеукраїнських та 1 регіональній конференціях: XII Науково-практична міжнародна конференція «Інформаційні

технології в освіті та управлінні», м. Нова Каховка, НКПІ, 2010 р.; XIII Науково-практична міжнародна конференція «Інформаційні технології в освіті та управлінні», м. Нова Каховка, НКПІ, 2011 р.; Четверта міжнародна науково-технічна конференція «Комп'ютерні системи та мережні технології» CSNT-2011, м. Київ, НАУ, 2011 р.; Всеукраїнська щорічна науково-методична конференція «Могилянські читання – 2011: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний», м. Миколаїв, ЧДУ ім. П. Могили, 2011 р.; Причорноморська регіональна науково-практична конференція професорсько-викладацького складу Миколаївського державного аграрного університету, м. Миколаїв, МДАУ, 2012 р.; Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» ISDMCI'2012, м. Євпаторія, 2012 р.; XIV Науково-практична міжнародна конференція «Інформаційні технології в освіті та управлінні», м. Нова Каховка, НКПІ, 2012 р.; П'ята міжнародна науково-технічна конференція «Комп'ютерні системи та мережні технології» CSNT-2012, м. Київ, НАУ, 2012 р.; Міжнародна науково-практична конференція молодих вчених та студентів «Студенти та молодь – для розвитку регіонів», м. Артемівськ, Навчально-науково-професійно-педагогічний інститут УПА, 2012 р.; Всеукраїнська щорічна науково-методична конференція «Могилянські читання – 2012: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний», м. Миколаїв, ЧДУ ім. П. Могили, 2012 р.; Третя міжнародна конференція молодих вчених та студентів «Сучасні інформаційні технології – 2013» MIT-2013, м. Одеса, ОНПУ, 2013 р.; VI Міжнародна науково-практична конференція молодих вчених та студентів «Інформаційні процеси та технології. Інформатика – 2013», м. Севастополь, СевНТУ, 2013 р.; Друга міжнародна науково-технічна конференція «Обчислювальний інтелект – 2013» (результати, проблеми, перспективи), ComInt-2013, м. Черкаси, ЧДТУ, 2013 р.; VII Міжнародна науково-практична конференція «Об'єктні системи – 2013», м. Ростов-на-Дону, Росія, 2013 р.; VII Міжнародна науково-практична конференція «Ольвійський форум – 2013: стратегії України в геополітичному просторі», м. Ялта, 2013 р.; Міжнародна науково-технічна конференція «Інформаційні технології та інформаційна безпека в науці, техніці та освіті «Інфотех – 2013», м. Севастополь, СевНТУ, 2013 р.; Наукова конференція «Фундаментальні та прикладні дослідження: інтеграція до світових наукометричних баз даних», м. Харків, 2013 р.; Всеукраїнська щорічна науково-методична конференція «Могилянські читання – 2013: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний», м. Миколаїв, ЧДУ ім. П. Могили, 2013 р.; XI Міжнародна науково-практичної конференція «Математичне та програмне забезпечення інтелектуальних систем», м. Дніпропетровськ, 2013 р.; Міжнародна наукова конференція «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту» ISDMCI'2014, смт Залізний Порт, 2014 р.; Всеукраїнська науково-методична конференція «Могилянські читання – 2014: досвід та тенденції розвитку

суспільства в Україні: глобальний, національний та регіональний», м. Миколаїв, ЧДУ ім. П. Могили, 2014 р.; Всеукраїнська науково-практична конференція молодих вчених, аспірантів і студентів «Інтелектуальні інформаційні системи», м. Миколаїв, ЧДУ ім. П. Могили, 2015 р.; IX Міжнародна науково-практична конференція «Ольвійський форум – 2015: стратегії України в геополітичному просторі», м. Миколаїв, 2015 р.; 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS); Всеукраїнська науково-методична конференція «Могилянські читання – 2015: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний», м. Миколаїв, ЧДУ ім. П. Могили, 2015 р.

Публікації. За темою дисертації було опубліковано 26 наукових робіт, з них 5 статей у журналах зі списку ВАК, 1 стаття у закордонному збірнику (Російська Федерація), 2 статті у збірниках, що входять до наукометричних баз (Ulrich's Periodicals Directory (США), DRIVER (ЄС), BASE (Німеччина), Index Copernicus (Польща), РИНЦ (Росія), ResearchBib (Японія)), 1 стаття у закордонному збірнику (Польща), що входить до наукометричної бази SCOPUS, 17 тез доповідей міжнародних, всеукраїнських та регіональних конференцій.

Структура дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків, списку літературних джерел та додатків. Загальний обсяг дисертації – 229 сторінок, 8 таблиць, 72 рисунки, 183 найменування використаних літературних першоджерел.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність напряму досліджень, наведено зв'язок роботи з науковими програмами, сформульовано мету і задачі дослідження, відображено наукову новизну, практичну цінність роботи й особистий внесок здобувача, наведені відомості про апробацію, публікації та використання результатів дослідження.

У **першому розділі** виконано аналіз існуючих методів інтеграції інформаційних технологій баз даних, OLAP та Data Mining. Показано, що побудова систем OLAP у середовищі об'єктних баз даних дозволяє вирішити проблему надмірності даних, що виникає при використанні ненормалізованих таблиць, і проблему збільшення кількості операцій при нормалізації бази даних, що призводить до зниження швидкості роботи системи в цілому.

При порівнянні реляційної моделі даних з об'єктною моделлю між ними можна знайти достатньо багато аналогій. Так, реляційне поняття таблиці відповідає класу, рядок таблиці – екземпляру класу, стовпець – властивості класу, зовнішній ключ – посиланню на об'єкт. Таким чином, таблиця фактів в об'єктній моделі даних являє собою клас, властивостями якого є міри та посилання на таблиці вимірів, які, в свою чергу, також є класами.

Відповідні зв'язки між собою мають класи інших вимірів. На відміну від таблиці фактів, яка містить посилання на екземпляри класів вимірів, класи

вимірів містять атрибути зв'язку з класами вимірів більш високого рівня для того, щоб не порушити цілісність бази даних.

При агрегуванні даних на більш високому рівні до вимірів можна звертатися за допомогою точкового синтаксису, застосовуваного в об'єктній моделі даних, наприклад:

Month.Quarter.CodeQuarter – код кварталу, в який входить місяць;

Month.Quarter.Year.CodeYear – рік, в який входить місяць (доступ через об'єкт класу «Квартал»).

Таким чином, стає можливим агрегування за вимірами з рівнями ієрархій великої кількості. При цьому немає надмірності даних, яка виникає при денормалізації таблиці фактів у MOLAP і схемі «зірка» в ROLAP. З іншого боку, агрегування за вищими рівнями ієрархії не буде таким витратним за часом, як при використанні схеми «сніжинка» в ROLAP.

Другий розділ присвячений математичній формалізації багатомірної моделі та методам створення OLAP-кубів у об'єктно-орієнтованому середовищі.

Множина-носіїв гіперкуба Ω може бути представлена у вигляді декартового добутку чотирьох представлених вище множин:

$$\Omega = D \times L \times A \times M,$$

де: D – множина вимірів гіперкуба для конкретної предметної області: $D = \{D_1, D_2, \dots, D_i, \dots, D_n\}$, де D_i – i -й вимір, n – кількість вимірів; L – множина всіх рівнів вимірів гіперкуба, яку можна представити у вигляді об'єднання множин рівнів кожного виміру гіперкуба: $L = L_1 \cup L_2 \cup \dots \cup L_i \cup \dots \cup L_n$; A – множина атрибутів (значень елементів) вимірів гіперкуба: $A = A_1 \cup A_2 \cup \dots \cup A_i \cup \dots \cup A_n$,

де A_i – множина атрибутів виміру D_i , яку в свою чергу можна представити як: $A_i = \{A_i^1, A_i^2, \dots, A_i^k, \dots, A_i^m\}$, де A_i^k – k -й атрибут i -го виміру, m – кількість атрибутів в i -му вимірі; M – множина значень мір гіперкуба: $M = \{M_{I_1, I_2, \dots, I_i, \dots, I_n}^1, \dots, M_{I_1, I_2, \dots, I_i, \dots, I_n}^l, \dots, M_{I_1, I_2, \dots, I_i, \dots, I_n}^z\}$, де I_i – індекс атрибуту i -го виміру, n – кількість вимірів, $M_{I_1, I_2, \dots, I_i, \dots, I_n}^l$ – l -та міра для комірки куба з індексом $I_1, I_2, \dots, I_i, \dots, I_n$, z – кількість мір гіперкуба.

Основним елементом багатомірної моделі даних є базовий куб, у якому зберігаються фактичні дані. Якщо до базового куба додати підсумкові або агреговані значення куба, які розглянуті нижче, отримується повний куб, побудова якого означає додавання до кожного з вимірів атрибуту зі значенням «ALL», що відповідає за підсумок за відповідним виміром.

Для ефективної побудови OLAP-системи в об'єктній СКБД необхідне використання розвинутої моделі метаданих. Для представлення метаданих найбільш логічним є використання тих самих структур, що й для фактичних даних: у реляційних БД вони зберігаються у таблицях спеціального призначення (системні таблиці), в об'єктних – як екземпляри класів, які для

даних випадків називаються метакласами. Запропонована модель метакласів представлена на рис. 1.

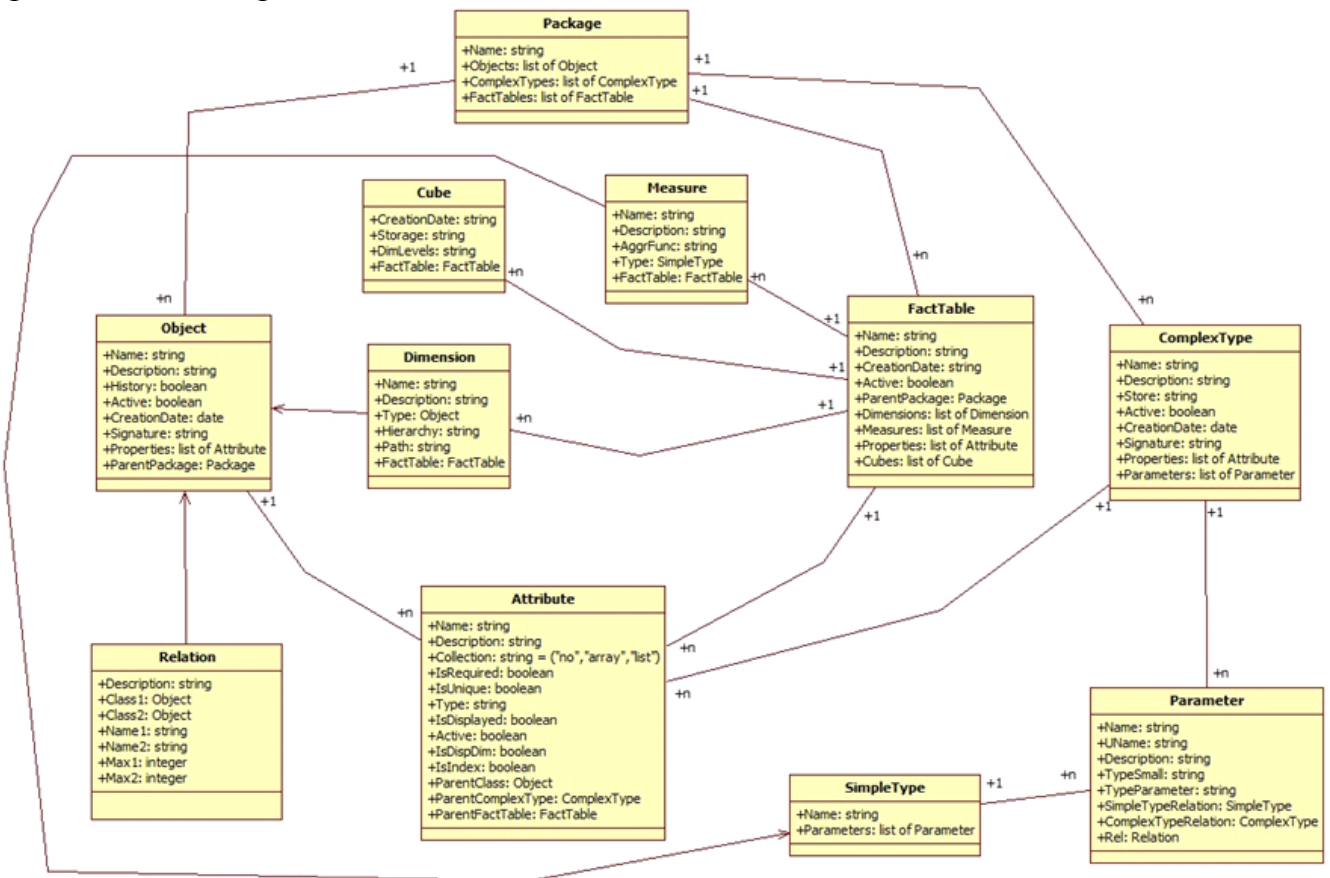


Рис. 1. Діаграма метакласів OLAP-системи в об'єктній СКБД

Таблиця фактів у системі метаданих по суті є різновидом звичайного класу, в якому атрибутами є виміри та міри, проте виміри завжди є об'єктними посиланнями, а міри – значеннями простого типу.

Будь-який куб можна представити деревом з кількістю рівнів, рівних кількості вимірів. Наприклад, існує тривимірний куб з вимірами $i, j, k - M_{ijk}$, де: $i = 1..p, j = 1..r, k = 1..s$. При представленні куба у вигляді дерева спочатку створюється коренева вершина, потім на 1-му рівні ієрархії дерева розміщуються p вершин дерева, що відповідають індексу i , на 2-му рівні – $p \times r$ вершин дерева, що відповідають індексам $|i, j|$, на 3-му рівні – $p \times r \times s$ $p \times r$ вершин дерева, що відповідають індексам $|i, j, k|$. Так, для $p = 2, r = 3$ і $s = 2$ таке дерево наведено на рис. 2.

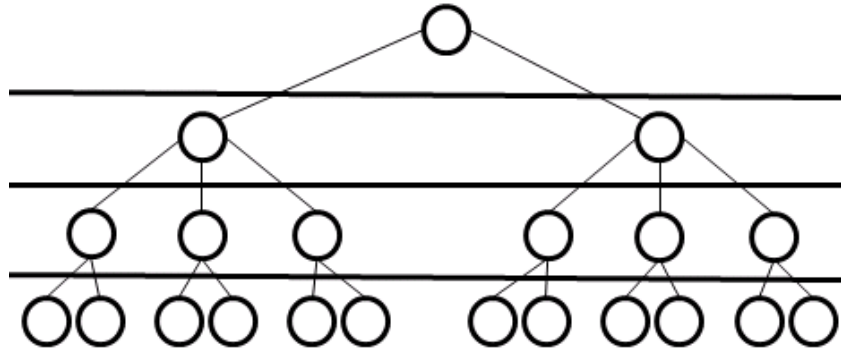


Рис. 2. Представлення куба у вигляді дерева

Пропонується алгоритм, за яким будується OLAP-куб у вигляді відповідного дерева. За даним алгоритмом проходження за фактичними даними відбувається тільки один раз, і для кожного об'єкта одночасно обраховується декілька агрегованих значень.

В основу алгоритму покладено той факт, що всі можливі агрегати за певним рівнем деталізації являють собою всі можливі комбінації за вимірами куба, кількість яких дорівнює C_n^k , де n – кількість вимірів, k – рівень деталізації.

Позначимо множину порядкових номерів вимірів куба як N , тобто $N = \{1, 2, 3, \dots, n\}$. У свою чергу, множини всіх можливих комбінацій елементів множини N за відповідним рівнем деталізації дорівнюють:

$$C^1 = \{\{1\}, \{2\}, \{3\}, \dots, \{n\}\};$$

$$C^2 = \{\{1,2\}, \{1,3\}, \dots, \{1,n\}, \{2,3\}, \dots, \{2,n\}, \dots, \{n-1,n\}\};$$

$$C^3 = \{\{1,2,3\}, \{1,2,4\}, \dots, \{n-2, n-1, n\}\}.$$

Сума всіх можливих підкубів дорівнює $C_n^0 + C_n^1 + C_n^2 + \dots + C_n^n = 2^n$. При цьому $C_n^0 = 1$ означає загальний агрегат за усім кубом, $C_n^n = 1$ – підкуб фактичних значень. Таким чином, кількість підкубів, в яких хоча б один вимір є фіксованим, а за всіма іншими вимірами здійснюється агрегування, дорівнюватиме $C_n^1 + C_n^2 + \dots + C_n^{n-1} = 2^n - 2$.

Суть алгоритму полягає у наступному: спочатку здійснюється генерація всіх можливих комбінацій множини N з рівнем деталізації від 1 до $n-1$, і отримані комбінації записуються у відповідний масив. Потім здійснюється проходження за всіма об'єктами класу фактів (далі – таблиця фактів), за яке один об'єкт почергово завантажується в оперативну пам'ять.

Наступним кроком є підрахунок загального підсумку куба. Відбувається це наступним чином: якщо був відкритий лише перший об'єкт, то, відповідно, створюється вершина:

$$V(\underbrace{0, 0, \dots, 0}_n),$$

в іншому випадку застосовується відповідна функція агрегування для збереженого раніше значення вершини та поточного значення факту:

$$V(\underbrace{0,0,\dots,0}_n) = f_{aggr}(V(\underbrace{0,0,\dots,0}_n), fact).$$

Всі інші можливі агрегати обраховуються за допомогою масиву *comb*, кількість елементів якого дорівнюватиме $2^n - 2$.

Запропонований комбінаторний алгоритм буде однаково підходити як для побудови куба з найбільш деталізованими даними, так і для консолідації даних. У другому випадку значення базового куба також є агрегованим, тому при виконанні запропонованого алгоритму при проходженні за кожним з фактів будуть знайдені такі факти, які мають однакові значення у всіх вимірах.

Третій розділ присвячено методам пошуку асоціативних правил у багатомірних даних як способу інтеграції технології OLAP і Data Mining та їх реалізації у середовищі об'єктних баз даних.

В основному методи Data Mining, включаючи і пошук асоціативних правил, працюють на даних, що представлені в табличному вигляді. Однак подібні закономірності можуть бути виявлені і в багатомірних даних.

Для полегшення розуміння процесу знаходження асоціативних правил між вимірами куба доцільно представити так звані їх шаблони, які будуть відрізнятися від правил тим, що в них замість конкретного атрибуту певного виміру вказується тільки відповідний вимір, їх у загальному вигляді можна записати наступним чином:

$$D_i \wedge \dots \wedge D_j = D_k \wedge \dots \wedge D_l,$$

де $D_i, D_j, D_k, D_l \subset D$ – певні виміри з множини вимірів, $i, j, k, l = 1 \dots n$ – певні порядкові номери вимірів куба, n – загальна кількість вимірів куба.

Цілком вірним буде твердження, що шаблон представляє деяку множину асоціативних правил, які йому відповідають.

Так, для тривимірного куба існують наступні шаблони асоціативних правил між ними.

Правила між двома вимірами:

$$D_1 \rightarrow D_2; D_2 \rightarrow D_1; D_1 \rightarrow D_3; D_3 \rightarrow D_1; D_2 \rightarrow D_3; D_3 \rightarrow D_2.$$

Правила між трьома вимірами:

$$D_1 \wedge D_2 \rightarrow D_3; D_3 \rightarrow D_1 \wedge D_2; D_1 \wedge D_3 \rightarrow D_2; D_2 \rightarrow D_1 \wedge D_3; D_2 \wedge D_3 \rightarrow D_1; D_1 \rightarrow D_2 \wedge D_3.$$

Чим більше стає вимірів у кубі, тим складніше знайти всі можливі шаблони асоціативних правил між ними.

У процесі знаходження асоціативних правил в OLAP-кубі можна виділити дві важливі стадії:

- генерація наборів даних, що часто зустрічаються і задовольняють визначеним обмеженням;
- власне генерація асоціативних правил.

На першому етапі у багатомірних даних можна знайти такі часті набори, які належать до абсолютно різних сукупностей. Нехай у загальному випадку множиною всіх частих предметних наборів у кубі OLAP є множина S :

$$S = \{S_1, S_2, \dots, S_i, \dots, S_n\},$$

де i – кількість елементів у предметному наборі, S_i – множина частих предметних наборів з кількістю елементів i , n – загальна кількість елементів у кубі.

У свою чергу:

- $S_1 = \{s_1, s_2, \dots, s_n\}$, де s_1 – множина частих одноелементних предметних наборів за першим виміром куба, s_2 – за другим виміром, s_n – за n -м виміром;
- $S_2 = \{s_{12}, s_{13}, \dots, s_{mn}\}$, де s_{12} – множина частих предметних наборів за сукупністю першого та другого вимірів, s_{13} – за сукупністю першого та третього вимірів, $m \neq n$.
- $S_k = \bigcup_{i=1}^{C_n^k} \{s_{i_1, i_2, \dots, i_k}\}$ – загальний вигляд представлення множини

k -елементних наборів.

При генерації частих предметних наборів у OLAP-кубі до них не потраплять всі елементи, що входять у відповідні виміри куба. Для того, щоб включити елемент або їх сукупність до таких наборів, спочатку необхідно обчислити підтримку даної сукупності.

Якщо ввести наступні позначення:

- $M_{\underbrace{i_1, i_2, \dots, i_t, \dots, i_n}_n}$ – значення міри у комірці куба, що відповідає значенням вимірів куба $i = \langle i_1, i_2, \dots, i_t, \dots, i_n \rangle$, $0 < i_t \leq n$, якщо t -й вимір куба має фіксоване значення, $i_t = 0$, якщо за t -м виміром відбувається агрегування;
- $M_{\underbrace{0, 0, \dots, 0}_n}$ – підсумкове значення всього куба;
- $ALL = \langle \underbrace{0, 0, \dots, 0}_n \rangle$ – сукупність значень вимірів, що відповідає значенню загального підсумку за кубом,

то підтримка багатомірного набору з куба дорівнюватиме:

$$Supp(i) = \frac{M_i}{M_{ALL}}.$$

Пропонується створювати частий предметний набір у вигляді списку, в якому перший елемент є підсписком, що містить порядкові номери вимірів кубу, за якими здійснюється генерація набору. Всі наступні елементи списку, тобто від другого до останнього, будуть містити інформацію про певний знайдений предметний набір.

У загальному випадку список має такий формат:

$$\begin{aligned} & \underbrace{\langle \langle id_1 \rangle, \langle id_2 \rangle, \dots, \langle id_k \rangle \rangle}_k, \\ & \underbrace{\langle \underbrace{\langle val_{11}, val_{21}, \dots, val_{k1} \rangle}_k \& Supp_1 \rangle, \dots, \langle \underbrace{\langle val_{1z}, val_{2z}, \dots, val_{kz} \rangle}_k \& Supp_z \rangle \rangle}_z \end{aligned}$$

де k – кількість елементів у предметному наборі, id_i – порядковий номер i -го виміру куба у відповідному предметному наборі, val_{ij} – значення атрибуту i -го виміру куба у відповідному j -му предметному наборі, $Supp_j$ – значення підтримки j -го предметного набору, z – отримана кількість частих предметних наборів.

Генерація частих предметних наборів складається з наступних етапів:

1. Генерація одноелементних наборів, при здійсненні якої потрібно перебрати всі значення атрибутів кожного виміру окремо один від одного.

2. Генерація двоелементних наборів, яка базується на основі двох одноелементних наборів, що представлені списками, які мають такий вигляд:

$$\begin{aligned} list_1 &= \langle \langle id_1 \rangle, \underbrace{\langle val_{11} \& Supp_1 \rangle, \dots, \langle val_{1z} \& Supp_z \rangle}_z \rangle; \\ list_2 &= \langle \langle id_2 \rangle, \underbrace{\langle val_{21} \& Supp_1 \rangle, \dots, \langle val_{2z} \& Supp_z \rangle}_z \rangle. \end{aligned}$$

Список частих двоелементних наборів отримується при застосуванні операції декартового добутку списків $list_1$ і $list_2$ та обчислення величин сформованих наборів:

$$itemset = \langle \langle id_1, id_2 \rangle, \langle val_{11}, val_{21} \& Supp_1 \rangle, \dots, \langle val_{1z}, val_{2z} \& Supp_z \rangle \rangle.$$

3. Генерація k -елементних наборів на основі двох $(k-1)$ -елементних наборів та одного $(k-2)$ -елементного, $k=3\dots n$.

При цьому необхідно, щоб два застосованих $(k-1)$ -елементних набори мали спільний вимір або їх сукупність, що міститься у сполучному $(k-2)$ -елементному наборі.

Наприклад, генерація триелементного набору здійснюється на основі двох двоелементних наборів, представлених списками $list_1$ та $list_2$ відповідно, та одним сполучним одноелементним набором, представленим списком $sublist$:

$$\begin{aligned} list_1 &= \langle \langle id_1, id_2 \rangle, \underbrace{\langle val_{a1}, val_{b1} \& Supp_1 \rangle, \dots, \langle val_{ax}, val_{bx} \& Supp_x \rangle}_x \rangle; \\ list_2 &= \langle \langle id_2, id_3 \rangle, \underbrace{\langle val_{b1}, val_{c1} \& Supp_1 \rangle, \dots, \langle val_{by}, val_{cy} \& Supp_y \rangle}_y \rangle; \\ sublist &= \langle \langle id_2 \rangle, \underbrace{\langle val_{b1} \& Supp_1 \rangle, \dots, \langle val_{bz} \& Supp_z \rangle}_z \rangle. \end{aligned}$$

4. Отримання загального списку частих предметних наборів, що має ієрархічну структуру:

$$MasOfSets = \langle \underbrace{Set_1, \dots, Set_i, \dots, Set_n}_n \rangle.$$

У свою чергу, кожний із списків Set_i розбивається на підсписки, кожен з яких містить часті предметні набори за певними вимірами. Кількість елементів у списку Set_i буде дорівнювати кількості сполучень вимірів за кількістю i елементів.

На другому етапі пошуку асоціативних правил на основі отриманого загального списку частих предметних наборів у багатомірних даних будуть згенеровані асоціативні правила. Представимо формули обчислень інших характеристик значущості міжвимірового асоціативного правила для загального випадку.

Введемо наступні позначення:

- $Ant = \{i_{ant_1}, i_{ant_2}, \dots, i_{ant_k}, \dots, i_{ant_n}\}$ – множина конкретних значень вимірів, що входить до умови асоціативного правила;
- $Cons = \{i_{cons_1}, i_{cons_2}, \dots, i_{cons_k}, \dots, i_{cons_n}\}$ – множина конкретних значень вимірів, що входить до наслідку асоціативного правила.

Тоді:

- достовірність: $Conf(i) = \frac{M_i}{M_{Ant}}$;

- ліфт: $Lift(R) = \frac{Conf(R)}{Supp(Cons)} = \frac{M_i \cdot M_{ALL}}{M_{Ant} \cdot M_{Cons}}$;

- леверидж:

$$Lev(R) = Supp(R) - Supp(Ant) \cdot Supp(Cons) = \frac{M_{All} \cdot M_i - M_{Ant} \cdot M_{Cons}}{M_{All}^2}.$$

Для задачі пошуку асоціативних правил у багатомірних даних до діаграми метакласів необхідно додати ще один метаклас, що представлятиме інформацію про міжвимірові асоціативні правила. Даний метаклас названо InterdimRule, він має зв'язок з метакласом Cube типу «один до багатьох», оскільки до одного куба може існувати кілька міжвимірових асоціативних правил. Даний метаклас містить у собі інформацію про класи, екземпляри яких являють собою асоціативні правила, що були отримані у багатомірних даних. На відміну від метакласу InterdimRule, такі класи належать до пакету, що містить класи сутностей певної бази даних, структура якої спроектована за допомогою метаданих, та містять властивості, які відображають інформацію про асоціативне правило: списки вимірів, що входять до його умови та наслідку, значення відповідних вимірів, а також характеристики значущості асоціативного правила.

У **четвертому розділі** описано розроблену інтелектуальну інформаційну систему, що була реалізована в результаті застосування запропонованих методів побудови OLAP-систем у середовищі об'єктних баз даних та методів генерації міжвимірових асоціативних правил у багатомірних даних.

Для забезпечення роботи системи з довільною БД незалежно від її предметної області в інформаційній системі реалізований засіб роботи з метаданими, який дозволяє проектування сутностей об'єктної БД з їх

подальшою компіляцією. Реалізовані операції створення нових сутностей (класів, складних типів, таблиць фактів) шляхом їх опису в базі метаданих, а також їх редагування та видалення.

Для забезпечення ефективного маніпулювання даними у певній БД, створеній відповідно до опису метаданих, реалізовано засіб ведення об'єктної БД, що автоматизує створення нових екземплярів відповідних класів, редагування, видалення, відновлення існуючих у БД екземплярів, а також розрив фактичних зв'язків між об'єктами.

У створеній і представленій в роботі інформаційній системі побудована підсистема OLAP, в якій реалізовано засіб створення таблиць фактів, а також додавання до них вимірів, мір та допоміжних атрибутів шляхом опису метаданих. У свою чергу, для побудови кубів OLAP у системі реалізовано відповідний механізм, який дозволяє побудувати OLAP-куби за допомогою запропонованого комбінаторного алгоритму. У межах підсистеми OLAP також реалізовано засіб візуалізації багатомірних кубів та побудови запитів до них.

Для забезпечення можливості інтелектуального аналізу кубів OLAP шляхом пошуку в них міжвимірових асоціативних правил реалізовано засіб їх генерації за допомогою описаних у третьому розділі методів та їх візуального представлення.

Розроблена інформаційна система складається з наступних модулів:

- модуль проектування БД (або модуль метаданих);
- модуль ведення БД (або модуль даних);
- модуль підсистеми OLAP;
- модуль інтелектуального аналізу.

Структуру інформаційної системи наведено на рис. 3.

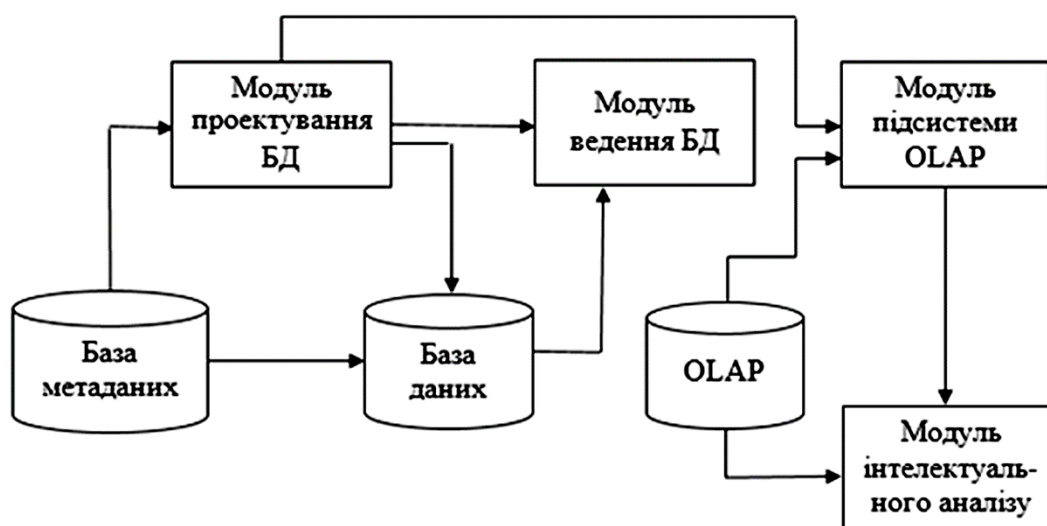


Рис. 3. Структура інтелектуальної інформаційної системи

Система реалізована у середовищі СКБД Caché за допомогою його серверу Caché Objects та є собою сукупністю взаємопов'язаних CSP-сторінок. Технологія CSP (Caché Server Page) відображає поточний стан БД і результати певних запитів шляхом генерації динамічних Web-сторінок.

Модуль проектування БД представляє засоби для створення класів, що відповідають сутностям БД, шляхом редагування бази метаданих.

Основними задачами модуля ведення БД є створення нового екземпляру певного класу (кінцевий користувач повинен обрати цей клас) з наступним збереженням у БД.

Модуль підсистеми OLAP призначений для вирішення задач оперативного аналізу даних, а саме:

- підготовка даних для здійснення над ними оперативного аналізу;
- здійснення оперативного аналізу шляхом формування багатомірного куба OLAP.

Останнім з модулів системи є модуль інтелектуального аналізу, який надає програмні засоби для пошуку асоціацій у OLAP-кубах.

Даний модуль надає можливість генерації наступних видів асоціативних правил у OLAP-кубах:

- повні міжвимірові асоціативні правила;
- неповні міжвимірові асоціативні правила;
- неповні міжвимірові асоціативні правила з фіксованим набором вимірів.

Отримані асоціативні правила є екземплярами створеного та скопійованого класу в СКБД Caché, опис якого є одним з екземплярів метакласу InterdimRule. Асоціативні правила зберігаються у БД не довгостроково, а тимчасово. Але у системі існує опція збереження асоціативних правил у текстовий файл, завдяки чому їх можна розглянути у будь-який момент часу незалежно від того, збережені вони на даний момент у БД чи вже були видалені.

У процесі даного дослідження була здійснена перевірка запропонованого методу генерації міжвимірових асоціативних правил для прикладу OLAP-куба БД деякої торговельної мережі. За допомогою представленої вище реалізованої інформаційної системи був побудований OLAP-куб, який має три виміри: «Дата» з ієрархією «Місяць», «Квартал», «Рік»; «Місце» з ієрархією «Магазин», «Регіон» та «Товар» з ієрархією «Товар», «Група товарів».

Експеримент проводився для різних ієрархій вимірів куба. Спочатку він був здійснений для куба з найбільш деталізованими даними (виміри мають рівні ієрархії «Місяць», «Магазин» та «Товар», відповідно), потім рівень ієрархії виміру «Дата» був підвищений до кварталу. Наостанок експеримент був здійснений для куба з найбільш узагальненими даними (рівні ієрархії «Рік», «Регіон» та «Група товарів»). Результати експерименту наведені у табл. 1.

Таблиця 1

Результати експерименту пошуку асоціативних правил у OLAP-кубах з різними рівнями ієрархій вимірів

№ з/п	Рівні ієрархій вимірів	Порогові значення		Знайдені асоціативні правила	
		Мінімальна підтримка	Мінімальна достовірність	З мінімальною достовірністю	З максимальною достовірністю

1	«Місяць», «Магазин», «Товар»	1%	30%	Якщо (Місяць="08-2014"), то (Товар="М'ясні продукти") Supp =1,59%; Conf =30,79%	Якщо (Товар="Молочні продукти" та (Місяць="08-2014"), то (Магазин="Тюльпан") Supp =1,15%; Conf =75,81%
2	«Квартал», «Магазин», «Товар»	1%	40%	Якщо (Магазин="Тюльпан") та (Квартал="3кв-2015"), то (Товар="М'ясні продукти") Supp =1,03%; Conf =40,55%	Якщо (Товар="Молочні продукти" та (Квартал="3кв-2014"), то (Магазин="Тюльпан") Supp =1,34%; Conf =54,69%
3	«Рік», «Регіон», «Група товарів»	10%	60%	Якщо (Група товарів="Продукти"), то (Регіон="Очаків") Supp =44,8%; Conf =60,15%	Якщо (Група товарів="Продукти") та (Рік=2014), то (Група товарів="Товари") Supp =44,8%; Conf =60,15%

Таким чином, при узагальненні багатомірних даних значення підтримок багатомірних предметних наборів та достовірностей міжвимірних асоціативних правил збільшується від найбільш деталізованих даних до найбільш узагальнених. Тому асоціації, що не є значимими у деталізованих даних, можуть виявитися значимими у випадку узагальнення даних. У будь-якому випадку розроблена інтелектуальна інформаційна система дозволяє знайти міжвимірні асоціативні правила у багатомірних кубах з будь-якими ступенями деталізації. Рівні ієрархії вимірів куба визначаються кінцевим користувачем при побудові куба або його редагуванні, а порогові значення мінімальних підтримок та достовірностей асоціативних правил, відповідно при їх пошуку.

ВИСНОВКИ

1. Доведено, що структурою об'єктної моделі даних, яка відповідає таблиці фактів у реляційній моделі даних, в об'єктній моделі даних є клас, властивостями якого є виміри та міри. При цьому міри звичайно є властивостями, що мають простий тип даних, у свою чергу як виміри є іншими класами у відповідній схемі даних. Таким чином, побудова системи OLAP для об'єктної моделі даних вирішує проблему надмірності даних, що звичайно виникає при використанні ненормалізованих таблиць, а також проблему збільшення кількості операцій при нормалізації бази даних.

2. Отримав подальший розвиток підхід до побудови структури метаданих для об'єктних баз даних шляхом її розширення класами підсистем OLAP та Data Mining, що полягає у застосуванні нових метакласів, які являють собою опис та структуру таблиць фактів, вимірів, мір та кубів у певній схемі даних та опис класів, екземплярами якого є асоціативні правила, що належать до

відповідного багатомірного куба. Запропонована структура метаданих для об'єктної СКБД дає змогу проектування структури OLAP-системи, що заснована на об'єктній моделі даних, з подальшим інтелектуальним аналізом багатомірних даних, збережених у системі.

3. Досліджені методи побудови багатомірних структур даних в об'єктно-орієнтованому середовищі, а саме – запропоновано підхід до побудови куба у вигляді дерева. Принцип побудови OLAP-куба за допомогою деревоподібної структури удосконалено шляхом запропонування методу, який використовує комбінаторні властивості агрегованих значень куба. Даний метод дозволяє з однаковою ефективністю здійснити повну матеріалізацію довільного багатомірного куба незалежно від рівнів ієрархій його вимірів. Також за допомогою запропонованого методу можна ефективно розрахувати як щільний, так і розріджений OLAP-куб.

4. Доведено, що у OLAP-кубах можуть мати місце різні типи асоціацій: міжвимірові, у межах одного виміру та гібридні; запропоновано наступну класифікацію асоціативних правил: повні, неповні та неповні з фіксованим набором вимірів.

5. Запропоновано представлення шаблонів міжвимірових асоціативних правил, що полегшує процес розуміння їх знаходження, а також запропоновано метод генерації даних шаблонів, який знаходить їх за допомогою генерації всіх можливих сполучень вимірів куба.

6. Розроблено метод генерації можливих асоціацій у багатомірних даних на базі отриманих відповідно до них шаблонів та властивості Apriori. Їх генерація здійснюється у два етапи: пошуку частих предметних наборів між різними вимірами в OLAP-кубі та генерації асоціативних правил.

7. Запропоновано метод формування частих предметних наборів у OLAP-кубах у вигляді багаторівневого списку та метод генерації міжвимірових асоціативних правил, виходячи з даного списку.

8. Розроблено інформаційну систему, що поєднує в собі інформаційні технології об'єктних БД, OLAP і Data Mining, яка являє собою Web-додаток, розроблений з використанням технології CSP для СКБД Caché.

9. Порівняно результати, що отримані за допомогою реалізованої інформаційної системи, з результатами, отриманими при використанні MS SQL Server 2008 R2. Виявлено, що у MS SQL Server поки немає готового інструменту для пошуку асоціативних правил у багатомірних даних, а алгоритм Microsoft Association Rules, що використовується у ньому, орієнтований на пошук асоціативних правил у реляційних даних.

10. Отримані у роботі наукові результати та програмні засоби впроваджені у навчальний процес кафедри інтелектуальних інформаційних систем Чорноморського державного університету імені Петра Могили та пройшли апробацію у спільному підприємстві «ГРАСКО» (м. Миколаїв).

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Фісун М. Т. Дослідження та реалізація синтаксичних алгоритмів створення OLAP-кубів / М. Т. Фісун, **Г. В. Горбань** // Вестник Херсонского национального технического университета. – 2010. – №2(38). – С. 110–117.
2. Фісун М. Т. Аналіз особливостей об'єктної та багатовимірної моделей даних в СКБД Caché / М. Т. Фісун, **Г. В. Горбань** // Вестник Херсонского национального технического университета. – 2011. – №2(41). – С. 116–124.
3. Фісун М. Т. Інтеграція даних мережевого трафіку мультисервісної корпоративної мережі з класами постріляційної СКБД Caché / М. Т. Фісун, І. М. Журавська, **Г. В. Горбань** // Наукові праці : науково-методичний журнал. – Вип. 161 (Т. 173) : Комп'ютерні технології. – Миколаїв : Вид-во ЧДУ ім. П. Могили, 2012. – С. 105–110.
4. Фісун М. Т. Дослідження реалізацій методів Data Mining в MS SQL Server 2008 R2 та порівняльний аналіз методів прогнозування з моделлю Триггера-Ліча / М. Т. Фісун, **Г. В. Горбань**, М. С. Лаушкіна // Вестник Херсонского национального технического университета. – 2012. – №2(44). – С. 137–146.
5. Фісун М. Т. Модели и методы построения системы OLAP для объектно-ориентированных баз данных / М. Т. Фісун, **Г. В. Горбань** // Информационные технологии и вычислительные системы. – М., 2014. – №1(209). – С. 41–45.
6. Фісун М. Т. Створення OLAP-кубів в об'єктно-орієнтованих СКБД за допомогою алгоритму DWARF / М. Т. Фісун, **Г. В. Горбань** // Наукові праці : науково-методичний журнал. – Вип. 201 (Т. 213) : Комп'ютерні технології. – Миколаїв: Вид-во ЧДУ ім. Петра Могили, 2013. – С. 99–105.
7. **Горбань Г. В.** Застосування V*-дерев для створення та обчислення OLAP-кубів з використанням комбінаторного алгоритму / **Г. В. Горбань** // Технологический аудит и резервы производства. – Х., 2013. – №5/4(13). – С. 10–12.
8. Фісун М. Т. Інтеграція технологій OLAP та Data Mining при побудові міжвимірних асоціативних правил / М. Т. Фісун, **Г. В. Горбань** // ScienceRise : науковий журнал. – 2015. – №6/2(11). – С. 103–111.
9. Fisun, M. Generation of Frequent Item Sets in Multidimensional Data by Means of Templates for Mining Inter-Dimensional Association Rules / M. Fisun, I. Kulakovska, **G. Gorban** // The 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. – 2015. – Vol. 1. – P. 368–376.
10. Фісун М. Т. Інтеграція даних щодо мережевого трафіку мультисервісної корпоративної мережі з класами постріляційної СКБД Caché / М. Т. Фісун, **Г. В. Горбань**, І. М. Журавська // Комп'ютерні системи та мережеві технології CSNT 2011 : тези доповідей IV Міжнародної

- науково-технічної конференції (14–16 червня 2011 р., м. Київ) ; Нац. авіац. ун-т. – К. : НАУ, 2011. – С. 63.
11. **Горбань Г. В.** Дослідження та аналіз моделей представлення даних на логічному рівні в СКБД Caché / **Г. В. Горбань** // Матеріали Причорноморської регіональної науково-практичної конференції професорсько-викладацького складу (18–20 квітня 2012 р., м. Миколаїв). – Миколаїв : МДАУ, 2012. – С. 46–48.
 12. Фісун М. Т. Побудова метаданих проектування системи оперативного і інтелектуального аналізу даних у об'єктних СКБД / М. Т. Фісун, **Г. В. Горбань** // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : матеріали міжнародної наукової конференції. – Херсон : ХНТУ, 2012. – С. 295–296.
 13. Фісун М. Т. Використання OLAP-технології для аналізу мережевого трафіку засобами об'єктної СКБД / М. Т. Фісун, І. М. Журавська, **Г. В. Горбань** // Комп'ютерні системи та мережні технології (CSNT-2012) : тези доповідей V Міжнародної науково-технічної конференції (13–15 червня 2012 р., м. Київ) / Національний авіаційний університет. – К. : НАУ, 2012. – С. 129–130.
 14. **Горбань Г. В.** Застосування об'єктної моделі даних при створенні системи оперативного та інтелектуального аналізу даних / **Г. В. Горбань** // Студенти та молодь – для розвитку регіонів : матеріали I Міжнародної студентської науково-практичної конференції молодих учених та студентів (18–19 жовтня 2012 р., м. Артемівськ) : в 4 т. – Артемівськ : ННПП УПА, 2012. – Т. IV : Інформаційні технології. Історія та культурознавство. – С. 17–19.
 15. **Горбань Г. В.** Методика поиска ассоциативных правил в многомерных данных / **Г. В. Горбань** // Інформаційні процеси і технології «Інформатика – 2013» : матеріали VI Міжнародної науково-практичної конференції молодих учених та студентів (22–26 квітня 2013 р., м. Севастополь) / Севастоп. нац. техн. ун-т ; наук. ред. С. В. Доценко. – Севастополь : Вебер, 2013. – С. 161–162.
 16. **Горбань Г. В.** Создание метаклассов системы анализа данных для объектно-ориентированных СУБД / **Г. В. Горбань** // Сучасні інформаційні технології 2013 : матеріали Третьої міжнародної конференції студентів і молодих науковців (25–26 квітня 2013 р., м. Одеса) / Одеськ. нац. політехнічний ун-т. – Одеса : ТЕС, 2013. – С. 100–101.
 17. Фісун М. Т. Модели и методы построения системы OLAP для объектно-ориентированных баз данных / М. Т. Фісун, **Г. В. Горбань** // Объектные системы – 2013 : материалы VII Международной научно-практической конференции (10–12 мая 2013 г., г. Ростов-на-Дону) / под общ. ред. П. П. Олейника. – Ростов н/Д : ШИ (ф) ЮРГТУ (НПИ), 2013. – С. 65–71.

18. **Горбань Г. В.** Пошук асоціативних правил в багатомірних даних як спосіб інтеграції технології OLAP та Data Mining / **Г. В. Горбань** // Обчислювальний інтелект (OI-2013) : матеріали Другої міжнародної науково-технічної конференції (14–17 травня 2013 р., м. Черкаси). – Черкаси, 2013. – С. 98–100.
19. Фісун М. Т. Підходи до створення систем OLAP в об'єктно-орієнтованих базах даних / М. Т. Фісун, **Г. В. Горбань** // «Ольвійський форум – 2013: стратегії країн Причорноморського регіону в геополітичному просторі» : тези. – Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2013. – Т. 7. – С. 274–280.
20. **Горбань Г. В.** Проектирование системы оперативного и интеллектуального анализа в объектных СУБД / **Г. В. Горбань** // Інформаційні технології та інформаційна безпека в науці, техніці та навчанні «ІНФОТЕХ-2013» : матеріали міжнародної науково-практичної конференції (9–13 вересня 2013 р., м. Севастополь) / Севастоп. нац. техн. ун-т ; наук. ред. О. В. Скатков. – Севастополь : СевНТУ, 2013. – С. 35–36.
21. **Горбань Г. В.** Генерація шаблонів асоціативних правил в OLAP-кубах / **Г. В. Горбань** // Математичне та програмне забезпечення інтелектуальних систем (MPZIS-2013) : XI міжнародна науково-практична конференція (20–22 листопада 2013 р., м. Дніпропетровськ) : тези допов. – Дніпропетровськ, 2013. – С. 61–63.
22. **Горбань Г. В.** Виявлення асоціацій у багатомірних даних шляхом знаходження шаблонів асоціативних правил / **Г. В. Горбань** // Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту : матеріали міжнародної наукової конференції. – Херсон : ХНТУ, 2014. – С. 265–267.
23. **Горбань Г. В.** Особливості пошуку асоціативних правил у багатомірних даних / **Г. В. Горбань** // Всеукраїнська науково-методична конференція «Могилянські читання – 2014 : досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти» : тези. – Миколаїв : Вид-во ЧДУ імені Петра Могили, 2014. – С. 58–62.
24. **Горбань Г. В.** Виявлення асоціацій у багатомірних даних як спосіб інтеграції інформаційних технологій OLAP та Data Mining / **Г. В. Горбань** // Всеукраїнська науково-практична конференція молодих учених, аспірантів і студентів «Інтелектуальні інформаційні системи» (16–17 лютого 2015 р, м. Миколаїв) : програма і тези. – Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2015. – С. 12–14.
25. **Горбань Г. В.** Характеристики значущості міжвимірних асоціативних правил у багатомірних даних та їх обчислення / **Г. В. Горбань** // «Ольвійський форум – 2015: стратегії країн Причорноморського регіону в геополітичному просторі» : тези. – Миколаїв : Вид-во ЧДУ ім. Петра Могили, 2015. – Т. 2. – С. 97–100.

26. **Горбань Г. В.** Пошук можливих асоціативних правил у багатомірних даних за допомогою генерації їх шаблонів / **Г. В. Горбань** // Всеукраїнська науково-методична конференція «Могилянські читання – 2015: досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти»: [збірник тез]. – Т. 1. – Миколаїв : Вид-во ЧДУ імені Петра Могили, 2015. – С. 75–77.

АНОТАЦІЇ

Горбань Г. В. Методи та об'єктно-орієнтована інформаційна технологія інтелектуального аналізу багатомірних даних. – На правах рукопису.

Дисертаційна робота на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Чорноморський державний університет імені Петра Могили МОН України, Миколаїв, 2016.

У дисертаційній роботі розв'язано актуальну науково-практичну задачу інтеграції моделей, методів та інформаційних технологій баз даних, OLAP та Data Mining у програмному середовищі об'єктних баз даних шляхом розроблення інструментальної програмної системи автоматизованого створення об'єктних багатомірних баз даних, методів формування асоціативних правил за шаблонами та їх реалізації в складі інструментальної системи.

Проаналізовано існуючі методи інтеграції інформаційних технологій баз даних, OLAP та Data Mining. Запропоновано принципи побудови OLAP-системи для об'єктних баз даних шляхом застосування моделі метаданих для її створення. Дана модель дозволяє здійснювати проектування довільної об'єктної бази даних на фізичному рівні і її безпосередній оперативний та інтелектуальний аналіз. Розглянуто побудову багатомірних структур даних в об'єктному середовищі та запропоновано підхід до побудови куба у вигляді деревоподібної структури, удосконалений шляхом запропонування алгоритму, який використовує комбінаторні властивості агрегованих значень куба. Досліджено можливі асоціації, що можуть бути знайдені у багатомірних даних, зокрема міжвимірні асоціативні правила, для полегшення процесу знаходження яких запропонована їх класифікація у відповідні шаблони. Досліджено застосування алгоритму Apriori на багатомірних даних та запропоновано його реалізацію. Розроблено інформаційну систему, що поєднує у собі інформаційні технології об'єктних БД, OLAP та Data Mining і представляє собою Web-додаток, розроблений з використанням технології CSP для СКБД Caché. Результати досліджень використані у виробничій діяльності підприємств м. Миколаєва.

Ключові слова: багатомірні бази даних, OLAP, вимір, міра, куб, рівень ієрархії, об'єктна модель даних, клас, метадані, рівень ієрархії, Data Mining, асоціативні правила, предметний набір, підтримка, достовірність.

Gorban G. V. Methods and object oriented information technology of multidimensional data mining. – Manuscript copyright.

Ph.D. thesis in Engineering Science on specialty No. 05.13.06 – Information technologies. – Petro Mohyla Black Sea State University of the Ministry of Education and Science of Ukraine, Nikolaev City, 2016.

The thesis is devoted to solving of the actual applied research task of integration of models, methods and information technologies, OLAP and Data Mining in the ODB software environment. It has been solved by development of the instrumental program system for computerized object multidimensional databases creation, by development of methods of template association rules forming and their implementation in the instrumental system.

Existing methods of integration of database information technologies, OLAP and Data Mining have been analyzed. Principles of OLAP-system building for object databases by means of metadata model application for its creation have been proposed. The model allows for physical design of the arbitrary object database and its direct operational analysis and data mining. Architecture of multidimensional data structures in the object environment has been reviewed and the approach for cube construction in tree format has been proposed. The approach was improved by means of the proposed algorithm which uses combinatorial properties of the aggregated cube data. Potential associations which can be found in multidimensional data have been investigated. In particular the association rules between dimensions have been investigated and their classification into respective templates has been proposed for easy finding of the association rules. Application of Apriori algorithm on multidimensional data has been investigated and realization of the algorithm was proposed. The information system combined with object DB, OLAP and Data Mining information technologies has been developed. The information system represents a Web application developed with the use of CSP technology for Caché DBMS. Results of the investigations have been used in business processes in Nikolaev City.

Key words: multidimensional databases, OLAP, dimension, measure, cube, level of a hierarchy, object data model, class, metadata, Data Mining, association rules, object set, support, validity.

Горбань Г. В. Методы и объектно-ориентированная информационная технология интеллектуального анализа многомерных данных. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – информационные технологии. – Черноморский государственный университет имени Петра Могилы МОН Украины, Николаев, 2016.

В диссертационной работе решена актуальная научно-практическая задача интеграции моделей, методов и информационных технологий баз данных, OLAP и Data Mining в программной среде объектных баз данных путем разработки инструментальной программной системы автоматизированного создания объектных многомерных баз данных, методов формирования ассоциативных правил по шаблонам и их реализации в составе инструментальной системы, которая позволяет с меньшими затратами

проектировать информационно-аналитические системы в объектной программной среде.

Проанализированы существующие методы интеграции информационных технологий баз данных, OLAP и Data Mining. Предложены принципы построения OLAP-системы для объектных баз данных путем применения модели метаданных для ее создания. Данная модель представляет собой совокупность классов, соответствующих основным понятиям объектной модели данных и сохраняющих информацию об их описании, и отношений между ними. В предлагаемой модели основным составляющим OLAP-систем в среде реляционных СУБД в объектной среде соответствуют классы таблиц фактов, измерений, мер и кубов. Модель позволяет осуществлять проектирование произвольной объектной базы данных на физическом уровне и ее непосредственный оперативный и интеллектуальный анализ.

Рассмотрено построение многомерных структур данных в объектной среде, предложен подход к построению куба в виде древовидной структуры и адаптирован известный алгоритм DWARF, что строит куб в названном виде. Усовершенствован механизм построения OLAP-куба в виде дерева путем разработки алгоритма, который использует комбинаторные свойства агрегированных значений куба. С помощью данного алгоритма можно построить как куб из самых детализированных данных, так и осуществить консолидацию данных путем повышения уровней иерархий соответствующих измерений куба.

Исследованы возможные ассоциации, которые могут быть найдены в многомерных данных, в частности ассоциативные правила между измерениями, для облегчения процесса нахождения которых предложена их классификация путем создания соответствующих шаблонов.

Предложен метод генерации шаблонов всех возможных ассоциаций между измерениями многомерного куба, позволяющий находить все шаблоны ассоциативных правил путем генерации соответствующих комбинаций измерений с их количеством от 1 до n , где n – общее число измерений. Приведен пример нахождения шаблонов ассоциативных правил в многомерных данных для OLAP-куба с четырьмя измерениями посредством использования описанного метода.

Разработан метод генерации возможных ассоциаций в многомерных данных на базе полученных в соответствии с ними шаблонов и свойства Apriori. Генерация ассоциативных правил между измерениями осуществляется в два этапа: на первом этапе находятся частые предметные наборы в данных OLAP-куба, а на втором на их основе генерируются соответствующие ассоциативные правила.

Показаны особенности поиска ассоциативных правил между измерениями в многомерных данных, а именно: в OLAP-кубе можно найти такие частые предметные наборы, которые относятся к совершенно разным совокупностям. Предложен метод формирования частых предметных наборов в виде многоуровневого списка. Также предложен метод генерации ассоциативных

правил между измерениями, исходя из данного списка, и выведены формулы расчета характеристик значимости ассоциативных правил для общего случая, когда есть куб с количеством N измерений.

Разработана информационная система, объединяющая в себе информационные технологии объектных БД, OLAP и Data Mining и представляющая собой Web-приложение, разработанное с использованием технологии CSP для СУБД Caché. Результаты исследований использованы в производственной деятельности предприятий г. Николаева.

Приведены примеры работы описанной информационной системы от этапа проектирования объектной БД на физическом уровне путем создания новых классов и связей между ними и наполнения их экстенгов объектами и до создания OLAP-куба с последующим выявлением в нем ассоциативных правил между измерениями.

Проведен эксперимент поиска ассоциативных правил между измерениями для трех OLAP-кубов с различными уровнями иерархий их измерений: наиболее детализированного, куба с повышенным уровнем иерархии в одном из измерений и куба с наиболее обобщенными данными. Выявлено, что в обобщенных данных могут иметь место ассоциации, которые не имеют смысла для детализированных данных, а также то, что пороговые значения минимальных поддержки и достоверности ассоциативных между измерениями правил целесообразно увеличивать при переходе от низших уровней иерархии измерений куба к высшим.

Проведено сравнение результатов работы реализованной интеллектуальной информационной системы и MS SQL Server 2008 R2 на примерах поиска ассоциативных правил, которое дает основание для вывода, что в MS SQL Server пока нет готового инструмента для поиска ассоциативных правил в многомерных данных.

Ключевые слова: многомерные базы данных, OLAP, измерение, мера, куб, уровень иерархии, объектная модель данных, класс, метаданные, Data Mining, ассоциативные правила, предметный набор, поддержка, достоверность.

Підп. до друку 10.03.2016. Формат 60×84/16.
Обл.-вид. арк. 0,9. Ум. друк. арк. 0,9.
Тираж 100 прим. Зам. 1003-1.

Поліграфічне підприємство СПД Румянцева Г. В.
54038, м. Миколаїв, вул. Бузника, 5/1.
Свідоцтво МК № 11 від 26.01.2007 р.