


МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ ПОЛІТЕХНІЧНИЙ УНІВЕРСИТЕТ

САФОНОВ Михайло Сергійович



УДК 004.042

МЕТОДИ ТА ТЕХНОЛОГІЇ УЗГОДЖЕННЯ ГЕТЕРОГЕННИХ ДАНИХ
В РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

05.13.06 – Інформаційні технології

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Одеса – 2015

Дисертацією є рукопис.

Робота виконана в Одеському національному політехнічному університеті
Міністерства освіти і науки України.

Науковий керівник

кандидат технічних наук, доцент
Яковенко Олександр Євгенович
Одеський національний
політехнічний університет,
кафедра природничо – наукової підготовки

Офіційні опоненти

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

доктор технічних наук, професор
Мещеряков Володимир Іванович
Одеський державний екологічний
університет,
завідувач кафедри інформатики

Захист відбудеться «03» липня 2015 р. о 13.30 на засіданні спеціалізованої
вченої ради Д 41.052.01 при Одеському національному політехнічному університеті
за адресою: 65044, м. Одеса, проспект Шевченка, 1, ауд. 204-А.

З дисертацією можна ознайомитись у бібліотеці Одеського національного полі-
технічного університету за адресою: м. Одеса, проспект Шевченка, 1.

Автореферат розіслано « » червня 2015 р.

Вчений секретар
спеціалізованої вченої ради



О. Є. Колесніков

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Сучасна тенденція до розподілених гетерогенних систем призвела до необхідності інтеграції різних типів даних. Більшість інформаційних систем (ІС) складаються з відокремлених автоматизованих систем (АС), які виконують визначені завдання і не виходять за межі своєї БД. БД можуть містити однакову інформацію та мати кілька різних архітектур, несумісних одна з одною, що вказує на їх гетерогенність. Існує задача забезпечення методів доступу до всіх необхідних даних з використанням розподілених БД, засобів реплікацій даних, управління подіями. Так як методи і підходи до зміни конкретної існуючої ІС не формалізовано, то засоби розробки програмних застосувань повинні задовольняти вимогам мобільності і, одночасно, роботи в гетерогенному середовищі розподіленої БД.

Інтеграція гетерогенних даних розглядаються переважно з позицій здійснення уніфікованих запитів, а не комплексного керування даними. Маніпулювання даними розглядається переважно при інтеграції тільки реляційних БД (або у випадку, якщо не реляційна БД має реляційну оболонку). У якості моделі даних розглядається реляційна схема, або в більш загальному випадку — табличне представлення даних, що значно збільшує час пошуку та перетворення гетерогенної інформації. Всі ці недоліки призводять до збільшення часу узгодження гетерогенних даних. В свою чергу часу узгодження включає в себе час пошуку, перетворення та передачі.

Істотне ускладнення інтеграції процесів і потоків даних в інформаційних системах, у результаті впровадження нових високорозвинених інформаційних технологій, і відсутність системи регулювання потоків інформації призводить до необхідності у знаходженні єдиного методу для роботи з різними компонентами даних і програмних застосувань, що скорочує час їх пошуку та перетворення. Метод повинен орієнтуватися на особливості всіх клієнтів в інформаційній системі та мати єдиний центр керування.

Таким чином, розробка клієнто-орієнтованого підходу перетворення гетерогенної інформації і технологій узгодження гетерогенних даних в розподілених інформаційних системах є **актуальною**.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана відповідно до пріоритетного наукового напрямку розвитку науки і техніки в Україні «Нові комп'ютерні засоби та технології інформатизації суспільства»; рішення Вченої ради ОНПУ від 27.10.14 р. пр. №3 щодо затвердження напрямків прикладної наукової діяльності у ХПТК ОНПУ; постанови КМУ від 13 липня 2011 р. N 752 «Про створення Єдиної державної електронної бази з питань освіти».

Мета та задачі дослідження. Мета дослідження полягає в скороченні часу пошуку та перетворення інформації, шляхом розробки методів та технологій узгодження гетерогенних даних в розподілених інформаційних системах з використанням клієнто-орієнтованого підходу .

Для реалізації основної мети були поставлені наступні наукові й практичні завдання:

- виконання аналізу в області розробки програмних систем узгодження гетерогенних даних, які використовують клієнт-серверні технології, та сумісності гетерогенних даних;

- розробка методу перетворення гетерогенної інформації;
- розробка моделі представлення інформації про стан виконання запитів в ІС;
- розробка інформаційної технології узгодження гетерогенних даних;
- впровадження розробленої інформаційної технології та експериментальне підтвердження скорочення часу пошуку та перетворення гетерогенних даних на прикладі інтеграції інформаційної системи з Єдиною державною електронною базою з питань освіти;

Об'єктом дослідження є процес узгодження даних в гетерогенній інформаційній системі.

Предметом дослідження є моделі та технології узгодження гетерогенних баз даних в інформаційній системі.

Методи досліджень. Основу дослідження становить клієнт-орієнтована технологія, що забезпечує взаємозв'язок між структурними елементами гетерогенної інформаційної системи та процесом пошуку і перетворення даних, і яка покладена в основу рефлекторно-комутаційного методу перетворення гетерогенної інформації. Використання теорії ймовірності та математичної статистики дозволило визначити емпіричну функцію розподілу часу виконань запитів програмними клієнтами та покладена в основу методу виявлення взаємозв'язаних списків. Математична логіка та теорія прийняття рішень використовувались для визначення правил поведінки програмних клієнтів в умовах гетерогенності.

При розробці програмного забезпечення системи пошуку та перетворення гетерогенних даних використовувалася технологія об'єктно-орієнтованого програмування.

Наукова новизна одержаних результатів:

- вперше розроблено рефлекторно-комутаційний метод перетворення гетерогенної інформації, відмінною особливістю якого є використання клієнто-орієнтованого підходу для передачі та узгодження гетерогенних даних, що дозволило скоротити час перетворення;
- вперше розроблено модель представлення інформації про стан виконання запитів в ІС на основі хромоматематичного підходу, що дозволило скоротити час прийняття оператором організаційно-технічних рішень;
- вдосконалено метод виявлення взаємозв'язаних списків, відмінною особливістю якого є розподіл виконання запитів на два етапи: спочатку складний запит розділяється на декілька простих і виконується локально на клієнтах, а завершальне формування загального результату — на сервері системи, що дозволило скоротити час пошуку гетерогенних даних в розподілених інформаційних системах;
- дістав подальшого розвитку метод узгодження гетерогенних даних в розподілених інформаційних системах шляхом використання клієнто-орієнтованого підходу, що дозволило підвищити ймовірність виконання запиту при скороченні часу на його виконання.

Практичне значення одержаних результатів. Розроблено математичне, алгоритмічне та програмне забезпечення узгодження гетерогенних даних в розподілених інформаційних системах. Побудована інформаційна технологія узгодження гетерогенних даних в розподілених інформаційних системах шляхом використання

клієнто-орієнтованого підходу, що дозволило підвищити ймовірність виконання запиту при скороченні часу на його виконання. Експериментальні дослідження дозволили підтвердити ефективність застосування клієнт-орієнтованої технології для оперативного отримання інформації від розподілених джерел інформації, що відповідає основним науковим напрямам, найважливішим проблемам фундаментальних досліджень у галузі природничих, технічних і гуманітарних наук на 2009–2015 роки та проекту концепції розвитку освіти України на період 2015 – 2025 років.

Результати дисертаційної роботи впроваджені в Херсонському політехнічному коледжі Одеського національного політехнічного університету, Київському технікумі електронних приладів та Миколаївському політехнічному коледжі. В результаті випробування встановлено, що впровадження інформаційної технології узгодження гетерогенних даних дозволило скоротити час перетворення гетерогенної інформації в середньому на 23%. Середній час сканування всіх БД та пошук потрібної інформації скоротився на 23,3%. Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованих систем трьох навчальних закладів дозволило в 8,7 разів скоротити час пошуку гетерогенної інформації. Оператори за допомогою моделі представлення інформації про стан виконання запитів в ІС витратили на прийняття організаційно-технічних рішень в середньому на 33,7% менше часу в порівнянні з табличним представленням даних.

Особистий внесок здобувача. Всі наукові й практичні результати отримані автором самостійно. В роботах, написаних спільно, автором отримані наступні результати:

в роботах [1, 12] – проведений аналіз існуючих методів узгодження даних на основі клієнто-орієнтованого підходу;

в роботах [2-3] – досліджені та сформульовані існуючі моделі узгодження даних в інформаційній системі;

в роботах [4, 6, 11] – представлено рефлекторно-комутаційний метод перетворення гетерогенних даних в інформаційній системі;

в роботі [5] – проведено моделювання процесу пошуку та перетворення даних в інформаційній системі;

в роботі [7] – проведений аналіз результатів впровадження та тестування системи пошуку та перетворення гетерогенних даних в існуючу інформаційну систему;

в роботах [8 - 10] – представлено метод виявлення взаємозв'язаних списків з розподілом виконання запитів на два етапи;

в роботах [13-16, 20] – розглянуто використання технології узгодження інформації в розподілених гетерогенних базах даних;

в роботах [17, 19] – представлено модель представлення інформації про стан виконання запитів в ІС на основі хромоматематичного підходу;

в роботі [18] – представлено математичну модель виявлення взаємозв'язаних списків для знаходження неявних зв'язків атрибутів;

в роботах [21, 22] – сформульовано процес побудови тривимірної моделі представлення інформації про стан виконання запитів;

в роботі [23] – представлено математичну модель рефлекторно-комутаційного методу перетворення інформації;

в роботі [24] – розглянуті розповсюдженні недоліки та проблеми впровадження інформаційної технології узгодження гетерогенних даних.

Апробація результатів дисертації. Основні висновки, положення і результати дисертації доповідалися і обговорювалися на Міжнародних науково-практичних конференціях: «Наукові дослідження та їх практичне застосування. Сучасний стан і шляхи розвитку» (Одеса, 2010, 2012); «Сучасні проблеми і шляхи їх рішення в науці, транспорті, виробництві і освіті» (Одеса, 2011, 2012); I Міжнародна конференція студентів і молодих учених. «Сучасні Інформаційні Технології» (Одеса, 2011, 2012); «Сучасні напрямки теоретичних та прикладних досліджень» (Одеса, 2012); I Міжнародна науково-практична інтернет-конференція «Проблеми та перспективи розвитку науки на початку третього тисячоліття у країнах СНД» (Переяслав-Хмельницький, 2012) ; друкувалися в збірниках наукових праць «Інформаційні технології в освіті, науці та виробництві» (Одеса, 2012, 2013); «Праці Одеського політехнічного університету. Комп'ютерні та інформаційні мережі та системи» (Одеса, 2013), II Україно-Німецька конференція "Інформатика. Культура. Техніка" (Одеса, 2014);

Публікації. За матеріалами дисертації опубліковано 24 друковані роботи, а саме: 7 у фахових виданнях, які входять до переліку ВАК України (з них 6 у виданнях, що входять до наукометричних баз Index Copernicus, Ulrich' s Periodicals Directory та eLibrary) та 17 статей і тез доповідей на науково-практичних конференціях.

Структура дисертації. Дисертація складається зі вступу, чотирьох розділів, висновків до кожного розділу, загальних висновків, списку використаних джерел (103 найменування, з них 36 іноземною мовою), 3 додатки. Робота містить 64 рисунка та 20 таблиць. Загальний обсяг дисертації – 142 друкованих сторінок, з них 130 сторінок основного тексту та 12 сторінок списку використаних джерел.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У *вступі* обґрунтована важливість і актуальність теми дисертації, викладені мета і завдання дослідження. Сформульовані основні положення і практичні результати, досягнуті в роботі, їх наукова новизна.

У *першому розділі* розглянуто проблеми незалежної модернізації окремих компонентів розподілених гетерогенних баз даних.

Аналіз шляхів розв'язку проблем фізичної гетерогенності, за допомогою введення стандартів взаємодії (ODBC, DAO, OLE DB, ADO, ADO.NET), показав, що перевагою даних механізмів є можливість застосування тих самих засобів доступу до різних типів джерел. Тому застосування легко модифікувати, якщо необхідна заміна СКБД. Але стандартизовані методи характеризуються неможливістю доступу до унікальної функціональності, специфічної для конкретної СКБД, та зниженням загальної продуктивності роботи з ними.

Серед проблем взаємозв'язку розподілених джерел даних в інформаційних системах, виділено проблеми: управління та передачі даних між джерелами з різною архітектурою, яку не вирішують існуючі методи; підвищення ефективності існуючої інформаційної системи без втручання в програмний код.

Розглянута можливість безпосереднього обміну інформацією між гетерогенними БД із застосуванням СКБД Oracle. Запити із БД Oracle до БД, що працюють під керуванням іншої СКБД, можуть здійснюватися за допомогою ODBC- драйверу й гетерогенного сервісу Oracle Heterogeneous Service. Недоліком такого методу є необхідність доробки автоматизованих систем, які обумовлені унікальністю БД, а уніфікація обміну між БД надскладна.

Важливим залишається питання віддаленого контролю виконання задач на рівні структурних підрозділів та створення моделі управління потоками даних в інформаційній системі, що представляють собою сукупність даних про суб'єкти освітньої діяльності, які функціонують в системі.

Постановка завдання дослідження формулюється як розробка нового методу перетворення гетерогенної інформації, що забезпечує скорочення часу пошуку та перетворення інформації в інформаційній системі та організації процесу автоматизованого збору гетерогенної інформації.

У *другому розділі* розроблено рефлексорно-комутаційний метод перетворення гетерогенної інформації в розподілених БД, що дозволяє скоротити час перетворення.

Основні принципи рефлексорно-комутаційного методу полягають у тому, що в будь-якій БД присутня певна множина сутностей

$$S = \{S_c | c = 1..D_c\}, \quad (1)$$

де D_c – загальна кількість сутностей в БД;

Кожна підмножина якої в свою чергу представляє собою:

$$S_c = \{S_{ci} | i = 1..m\} \quad (2)$$

де S_{ci} – кортеж даних сутності S_c ,

m – загальна кількість кортежів в сутності S_c ;

$$S_{ci} = [S_{i1}, S_{i2}, \dots, S_{in}], i = \overline{1, n} \quad (3)$$

n – загальна кількість атрибутів в сутності S_c ;

Аналіз запитів, що виконувались, показав що в основній більшості випадків час обробки (T_{bd}) збільшується пропорційно їхній кількості й дорівнює сумі часу, витраченого на обробку кожної бази:

$$T_{bd} = \sum_{i=1}^n T_i \quad (4)$$

Для більш точного визначення часових характеристик було введено час підключення до визначеної сутності (t'_c) і час, що потрібно для одержання відповіді (t''_c).

В ході дослідження виявлено, що найчастіше БД розміщені на окремих ПК. Тому потрібно розподілити частину часу обробки на виділені ПК. Одним з розв'язків даної проблеми є використання спеціалізованих програмних клієнтів, які брали б на себе операції перебору, а також прийому й передачі інформації головній системі управління (ГСУ). В якості клієнта системи треба розглядати програмний компонент, який буде здійснювати функцію комунікації з ГСУ.

Всі інструкції для потоків інформації повинні зберігатись в ГСУ та дублюватись для кожної підсистеми управління.

Нехай для кожного з G клієнтів існує своя множина інструкцій $P = \{p_i\}$, де $i = \overline{1, d_i}$, d_i – загальна кількість інструкцій для i -го клієнта. Згідно з кожним набором інструкцій для всіх клієнтів формується потік інформації I_c , $c \in (1, G)$.

Управління потоками відбувається централізовано. В ГСУ формуються інструкції, які відправляються на віддалені станції. Для зберігання інструкцій та запобігання помилок, викликаних розривом зв'язку, кожний клієнт має свою власну базу знань. Структура рефлексорно-комутаційного методу перетворення гетерогенної інформації в розподілених БД представлена на рисунку 1.

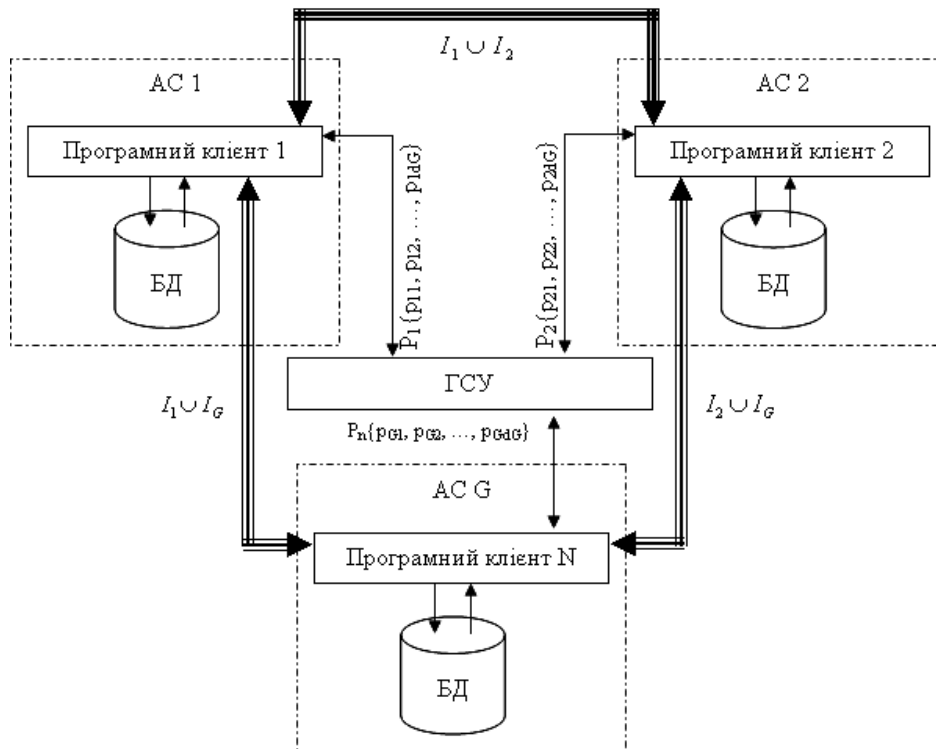


Рисунок 1 – Структура рефлексорно-комутаційного методу перетворення гетерогенної інформації в розподілених БД

В рамках рефлексорно-комутаційного методу перетворення гетерогенної інформації для формування набору інструкцій та отримання результатів у вигляді записів використовується SQL-орієнтована мова. Після відправки команди, ГСУ відправляє наступну іншому програмному клієнту і так далі. Після відсилання всіх команд відбувається збір інформації з тих же робочих станцій. Такий метод значно скорочує час T повної обробки всіх станцій, так як обробка команд клієнтами відбувається майже одночасно, а не послідовно.

Не виключена ситуація, коли та або інша робоча станція буде не готова надати відповідь у призначений строк. І тоді ГСУ прийдеться очікувати якийсь час Δt своєї бездії. У такий спосіб відбувається збільшення часу T_{bd} на деяку величину Δt . І завдання полягає в тім, щоб максимально зменшити її значення ($\Delta t \rightarrow 0$).

$$T_{bd} = \sum_{i=1}^n (t_i + t'_i + t''_i) + \Delta t \quad (5)$$

У зв'язку із цим доцільно ввести коефіцієнт λ , що буде відображати пріоритет і послідовність відсилання команд й одержань відповідей.

$$\lambda_1 < \lambda_2 < \dots < \lambda_G \Rightarrow t_1 > t_2 > \dots > t_G \quad (6)$$

Загальний час затримки буде залежати, більшою мірою, від часу роботи програмного клієнта з найвищим пріоритетом, і у меншому ступені від часу відправлення команди й прийому відповіді. Тому виведемо формулу визначення загального часу обробки всіх сутностей з урахуванням коефіцієнта:

$$T_{bd} = t'_1 + t''_G + \max(t_i), i \in [1..G] \quad (7)$$

У випадку, коли потрібно виконати w дій U_{ij} ($i \in [1..q], j \in [1..w]$) в q сутностях, при чому результат кожної дії повинен отримуватися ГСУ, необхідно організувати циклічний запит. Тобто обробка проходить від першої сутності до сутності q , а потім знову до першої і так далі поки всі дії не будуть виконані (рис. 2).

$$U_{11} \rightarrow U_{21} \rightarrow \dots \rightarrow U_{q1} \rightarrow U_{12} \rightarrow U_{22} \rightarrow \dots \rightarrow U_{q2} \rightarrow \dots \rightarrow U_{qw} \quad (8)$$

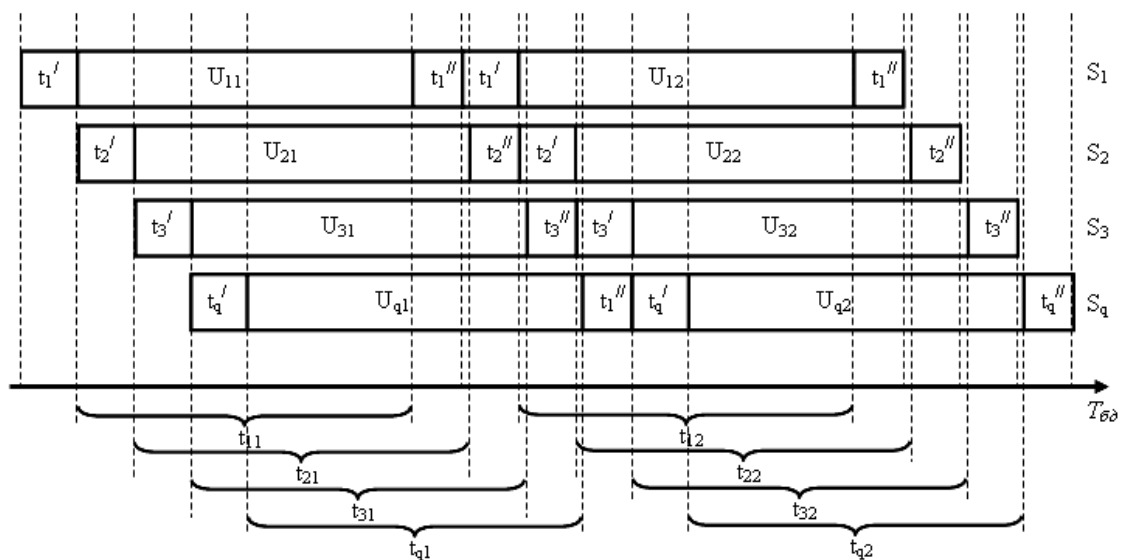


Рисунок 2 – Часова діаграма рефлекторно-комутаційного методу перетворення гетерогенної інформації з використанням програмних клієнтів

Даний метод спрощує та скорочує час передачі даних між ГСУ та програмними клієнтами за рахунок виділення часу обробки на визначені робочі станції. Згідно з рефлекторно-комутаційним методом загальний час перетворення даних залежить від максимального часу перетворення однієї БД, на відміну від класичного методу, в якому час дорівнює сумі часу перетворень даних в кожній БД.

Оператор від ГСУ, в межах рефлекторно-комутаційного методу, отримує такі дані: часу з'єднання з БД, час виконання складного запиту та час переходу від одного програмного клієнта до іншого. На сьогоднішній час оператор отримує всі дані у вигляді зведених таблиць, що призводить до збільшення часу прийняття оператором організаційно-технічних рішень.

В межах експерименту був оцінений час прийняття таких рішень для 10 програмних клієнтів. Всі клієнти виконували певні запити в середньому кожні 20 хвилин на протязі 8 годин в день. Експеримент тривав тиждень. За цей час

$$\begin{cases} x_i' = -\frac{T_{Ri} \cdot \pi}{G} \cdot \sin\left(\frac{\pi \cdot i}{G}\right) \\ y_i' = \frac{T_{Ri} \cdot \pi}{G} \cdot \cos\left(\frac{\pi \cdot i}{G}\right) \\ z_i' = \frac{T_{bd}}{2 \cdot w \cdot G} \end{cases} \quad (9)$$

Якщо $i \in [0, G]$, то $\frac{\pi \cdot i}{G} \in [0, \pi]$. Визначимо довжину спірального шляху L .

$$L = \int_0^\pi \sqrt{\left(-\frac{T_{Ri} \cdot \pi}{G} \cdot \sin\left(\frac{\pi \cdot i}{G}\right)\right)^2 + \left(\frac{T_{Ri} \cdot \pi}{G} \cdot \cos\left(\frac{\pi \cdot i}{G}\right)\right)^2 + \left(\frac{T_{bd}}{2 \cdot w \cdot G}\right)^2} d\left(\frac{\pi \cdot i}{G}\right) \quad (10)$$

Спростимо дане рівняння:

$$L = \left(\left(\frac{T_{Ri} \cdot \pi}{G}\right)^2 + \left(\frac{T_{bd}}{2 \cdot w \cdot G}\right)^2 \right) \cdot a \int_0^\pi = \frac{\pi}{G} \sqrt{(T_{Ri} \cdot \pi)^2 + \left(\frac{T_{bd}}{2 \cdot w}\right)^2} \quad (11)$$

Представлене рівняння (11) дозволяє розрахувати загальну довжину спірального шляху L , який характеризує часові параметри роботи рефлекторно-комутаційного методу перетворення гетерогенної інформації з використанням програмних клієнтів.

Довжина спірального шляху здебільшого залежить від часу обробки запиту на робочій станції клієнтом (ti') та часу переходу управління від однієї станції до іншої (Δti). Саме введення пріоритетів роботи системи вплине на часові характеристики роботи системи. Тобто, в процесі формування моделі будуть визначатися участки з часовими характеристиками, що значно більші за інші (відповідно будуть зображатися іншим кольором). Це буде означати, про доцільність зміни поточного пріоритету роботи програмного клієнта. За допомогою рівняння (10) можливо оперативно визначати характеристики стану інформаційної системи та своєчасно виявити причини затримки потоків даних, за рахунок ведення накопичуваної бази знань попередніх результатів роботи методу.

На основі моделі представлення інформації про стан виконання запитів в ІС оператори, в рамках описаного вище експерименту, витратили на прийняття організаційно-технічних рішень в середньому на 33,7% менше в порівнянні з табличним представленням.

В *третьому розділі* описаний метод виявлення взаємозв'язаних списків, який дозволив скоротити час виконання складних запитів до розподілених гетерогенних БД.

В різних сутностях може бути відображена різна інформація про один і той самий суб'єкт. На сучасному етапі скануються всі дані атрибутів баз даних для організації процесу реплікації даних. Автором запропонований метод виявлення взаємозв'язаних списків, згідно якому скануються тільки зв'язані списки, за рахунок чого скорочується час виконання складних запитів до розподілених гетерогенних БД.

Якщо існують деякі сутності S_1 та S_2 , які зберігають інформацію про m суб'єктів. Між ними існують неявні зв'язки $L_i, i \in [1..m]$, які визначають різні знання про один і той самий суб'єкт в різних сутностях. Неявний зв'язок – зв'язок двох

таблиць, що задається не в SQL-запиті, а шляхом визначення атрибутів, які зберігають схожі набори даних.

Припустимо сутність S_i складається з n значень атрибутів:

$$S_{ci} = [s_1, s_2, s_3, \dots, s_n] \quad (12)$$

Кожний елемент множини є носієм знань про визначений суб'єкт сутності. Залежність між атрибутами визначена певною функцією $f(s_i, s_j)$.

Визначимо множину F ймовірностей існування залежностей між атрибутами:

$$F = \{\mu_F(s_{c1})/s_{c1}, \mu_F(s_{c2})/s_{c2}, \dots, \mu_F(s_{cn})/s_{cn}\}, \quad (13)$$

де $\mu_F(s_i) \in [0..1]$ — оцінка статистичної ймовірності s_i залежності.

Операції об'єднання множин відповідає операція \max , що виконується над їх функціями приналежності:

$$\mu_D(s_{cj}) = \max(\mu_{Fi}(s_{cj})), i \in [1..h], j \in [1..n] \quad (14)$$

В умовах існування сутностей коефіцієнт μ_F визначається як статистична приналежність визначеним умовам. Для знаходження функції $f(S_1, S_2)$ потрібно визначити матрицю залежностей значень атрибутів обраних сутностей. Залежність між двома значеннями атрибутів визначимо як статистичну величину відношення між значеннями $s_{j1} \rightarrow s_{j2}$, $j \in [1, h]$.

На прикладі двох сутностей S_1 та S_2 , апроксимуємо функцію $f(S_1, S_2)$, та представимо її у вигляді об'єднання декількох простих функціональних залежностей:

$$\begin{aligned} f(S_1, S_2) \rightarrow & f_1(s_{11}, s_{12}) \vee f_2(s_{11}, s_{22}) \vee \dots \vee f_h(s_{11}, s_{h2}) \vee f_{h+1}(s_{21}, s_{12}) \vee \\ & \vee f_{h+2}(s_{21}, s_{22}) \vee \dots \vee f_{2-h}(s_{21}, s_{h2}) \vee \dots \vee f_{h-h}(s_{h1}, s_{h2}) \end{aligned} \quad (15)$$

Визначимо функціональні залежності для значень $s_{11} \rightarrow s_{12}, s_{22}, \dots, s_{h2}$, тобто значення $f_1(s_{11}, s_{12})$. Для цього розділимо всі значення s_{j1} на z груп в залежності від якісного значення.

Кожній із груп відповідає умова критеріального відбору $k_{li}, i \in [1, z], l \in [1, z]$. При чому для кожної сутності існує свій критерій відбору. Наступний розподіл значень атрибутів в сутностях за групами відбувається з урахуванням першого розподілу. Перший розподіл відбувається з урахуванням $k_{li}, i \in [1, z]$. Тобто всі значення першого атрибуту s_l розподіляються на групи у відповідності з критерієм відбору.

Тобто формуються множини значень атрибутів, де всі елементи множини S_{li} відповідають критерію k_{li} . Другий атрибут s_2 зв'язаний з першим, як носій інформації про один і той самий суб'єкт аналізу: студент, співробітник, тощо, або за ключем, якщо ці атрибути знаходяться в межах однієї БД. Атрибут s_2 також розподіляється за групами у відповідності до $k_{2i}, i \in [1, z]$.

Основним етапом при проектуванні системи пошуку та перетворення гетерогенної інформації, є розділення майбутньої системи на автономні структурно-логічні частини, взаємодія яких буде реалізовувати найбільшу функціональність і забезпечувати процес пошуку та перетворення: підсистема «Клієнт», підсистема «ГСУ» та підсистема «Аналіз стану інформаційної системи» (рис. 4).

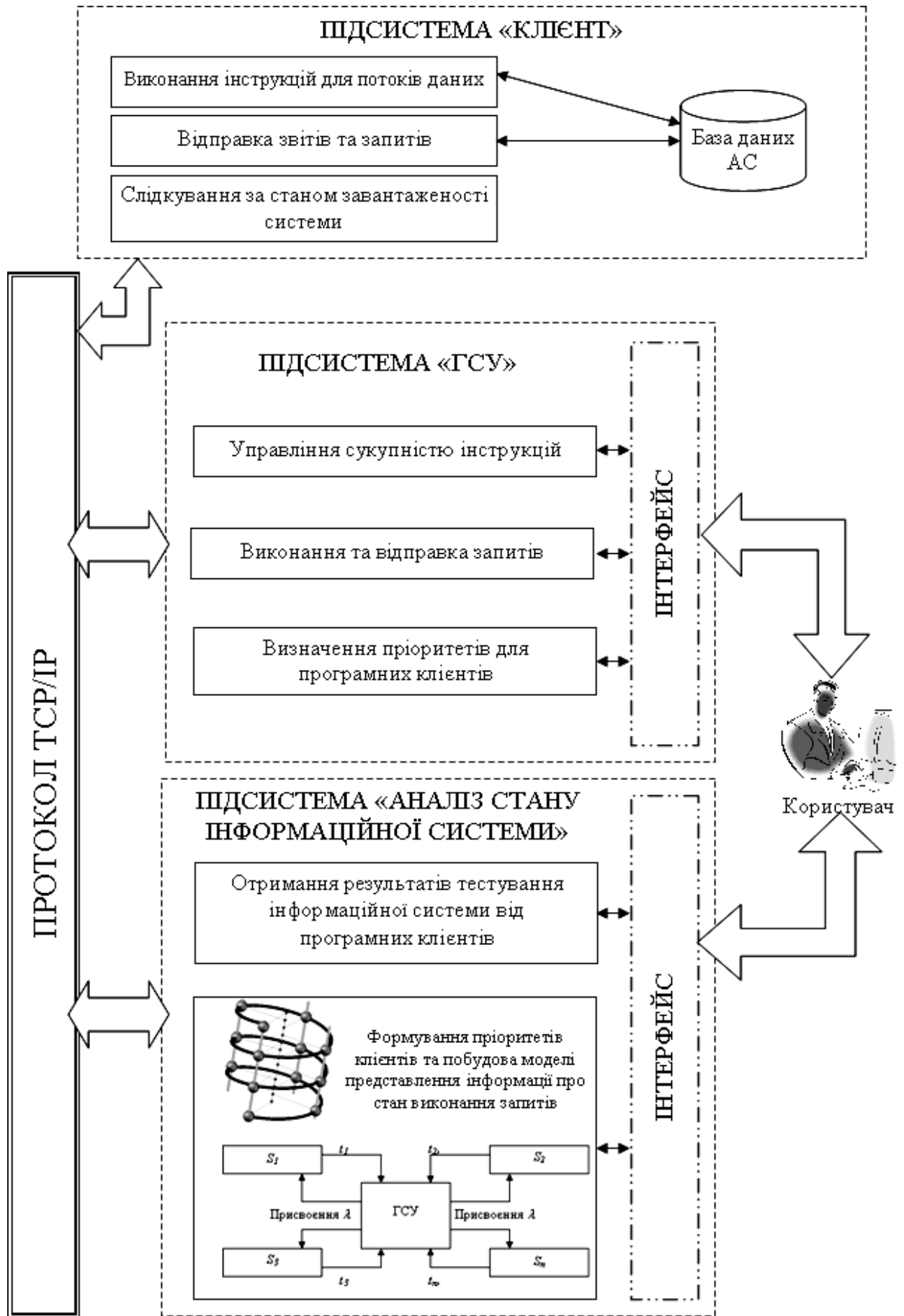


Рисунок 4 – Схема взаємодії програмних підсистем

Всі програмні клієнти надсилають звіти до ГСУ, яка й виконує збір та обробку отриманої інформації та побудову моделі. Згідно надісланих звітів формуються рейтинги клієнтів. Рейтинг визначає порядок та час виконання станціями визначених задач. Для розподілу завдань робочим станціям присвоюється коефіцієнт, що відображає пріоритет і послідовність відсилання команд й одержаних відповідей.

З часом стан інформаційної системи змінюється. Тому тестування станцій програмними клієнтами повторюється через визначені оператором проміжки часу. В подальшому кількість витків збільшується відповідно до збільшення кількості запитів на тестування клієнтами.

В якості вихідних даних підсистема може сформувати прогресивну функцію стану інформаційної системи. Тобто за допомогою прогностичних методів та знань про минулі стани системи існує можливість взнати майбутні стани системи з певною ймовірністю.

Програмне забезпечення клієнтів складається таким чином, щоб аналізувати роботу мережевого адаптера і декодувати одержувані дані, та додаткового програмного коду, що залежить від типу топології досліджуваної інформаційної системи. Крім того, поставляється ряд процедур декодування, орієнтованих на певний протокол, наприклад, IPX.

До складу програмних клієнтів входить підсистема аналізу, яка пропонує користувачеві рекомендації щодо дій, які слід проводити в певній ситуації, значень тих чи інших результатів вимірювань, усунення деяких видів несправності.

Одним з етапів інтеграції ЄДЕБО з існуючою АС, в межах рефлекторно-комутаційного методу перетворення гетерогенної інформації, було застосування методу виявлення взаємозв'язаних списків. Після впровадження даного методу був проведений експеримент, який полягав в імпорті списку всіх студентів по групам першого курсу. Імпорт відбувався за допомогою можливостей експорту ЄДЕБО, з подальшим копіюванням у власну АС, та за допомогою методу виявлення взаємозв'язаних списків.

Всього організовувалось копіювання 263 студентів з 10 груп. Пошук даних за допомогою можливостей експорту ЄДЕБО відбулося за 1390 секунд. З використанням методу виявлення взаємозв'язаних списків процес відбувся за 166 секунди. На перенесення 263 студентів за допомогою запропонованого методу було витрачено у 8,4 рази менше часу, ніж класичним методом, що каже про доцільність використання методу виявлення взаємозв'язаних списків у межах рефлекторно-комутаційного методу перетворення даних. Ймовірність знаходження хибного зв'язаного списку склала 0,03.

У **четвертому розділі** розроблена інформаційна технологія узгодження гетерогенних даних.

Дана технологія включає в себе використання протоколу TCP/IP для взаємозв'язку підсистем «Клієнт», «ГСУ» та «Аналіз стану інформаційної системи».

Алгоритм рефлекторно-комутаційного методу перетворення гетерогенної інформації з використанням програмних клієнтів представлений на рис. 5.

Підсистеми «ГСУ» та «Клієнт» мають власні бази знань, в яких формуються знання про потоки інформації, систематизуються та зберігаються в базі знань. База знань формується в процесі аналізу надходження даних від клієнтів та в результаті

виконання інструкцій. Усі описані вище підсистеми повинні працювати за схемою «Клієнт-Сервер». Взаємодія «клієнтів» з «сервером» повинна здійснюватися за допомогою протоколу TCP/IP, тому що він реалізує відправлення адресату послідовності сегментів, які утворюють деякий блок даних.

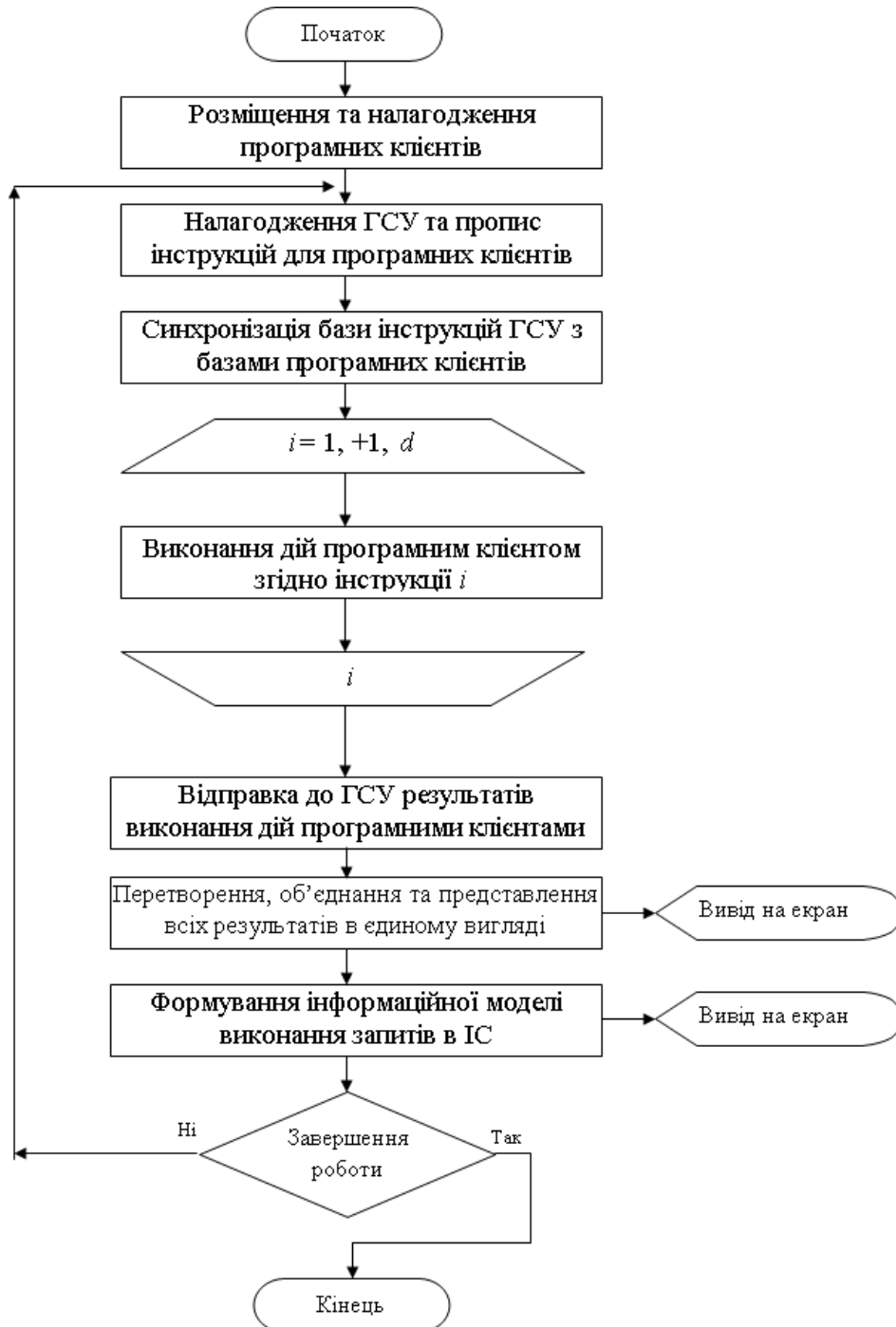


Рисунок 5 – Алгоритм рефлекторно-комутаційного методу перетворення гетерогенної інформації з використанням програмних клієнтів

Алгоритм методу виявлення взаємозв'язаних списків представлений на рис. 6.

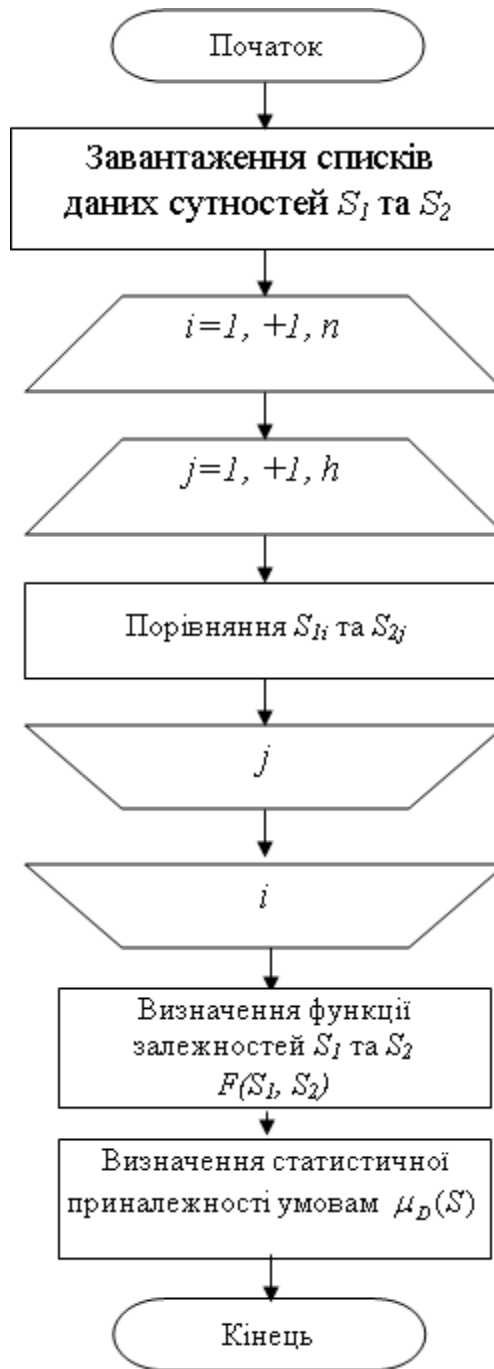


Рисунок 6 – Алгоритм методу виявлення взаємозв'язаних списків

На основі рефлекторно-комутаційного методу перетворення інформації та методу виявлення взаємозв'язаних списків була побудована інформаційна технологія узгодження гетерогенних даних яка представлена на рис. 7.

З метою перевірки ефективності запропонованих моделі та методів був проведений експеримент, який полягав у впровадженні в Херсонському політехнічному коледжі Одеського національного політехнічного університету (ХПТК ОНПУ), Київському технікумі електронних приладів (КТЕП) та в Миколаївському політехнічному коледжі (МПТК) системи пошуку та перетворення гетерогенної інформації.

Під час проведення експерименту кожна робоча станція була підключена до однієї мережі з сервером, на якому встановлена ГСУ.

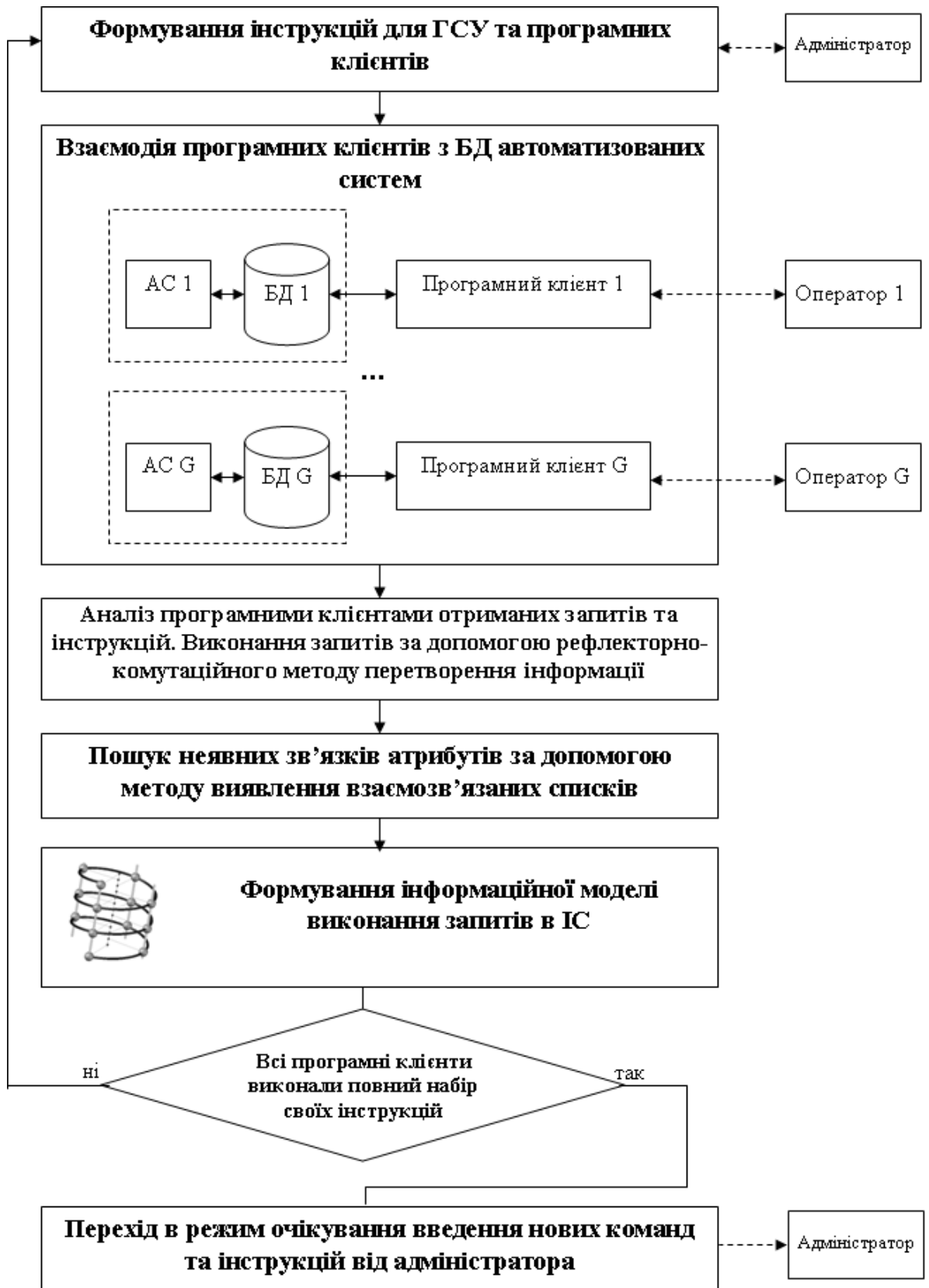


Рисунок 7 – Інформаційна технологія узгодження гетерогенних даних з використанням рефлексорно-комутаційного методу перетворення інформації та методу виявлення взаємозв'язаних списків

До ГСУ відправлялись від оператора складні SQL-запити, які потребували доступу до декількох БД, розміщених на різних робочих станціях, одночасно. Для експерименту загальна кількість запитів була обрана рівною десяти і відправлялись вони з інтервалом в 5 секунд. ГСУ поділяла складні запити на кілька простих і відправляла їх відповідним клієнтам. Які в свою чергу передавали їх безпосередньо до БД. Після виконання запитів, результати отримувались програмними клієнтами та передавались назад до ГСУ, де об'єднувались згідно початкового складного запиту. Схема розділу складного запиту на декілька простих представлена на рис. 8.



Рисунок 8 – Схема розділу складного запиту на декілька простих

В результаті впровадження технології узгодження гетерогенних даних в інформаційні системи ХПТК ОНПУ, була підвищена ефективність роботи локальних АС, за рахунок автоматизованої актуалізації даних. В результаті розробки алгоритмічного і програмного забезпечення компонентів інформаційної системи, це дозволило скоротити час перетворення гетерогенної інформації на 18% (рис. 9).

Час сканування всіх БД та пошук потрібної інформації скоротився на 11% за рахунок розділення складного запиту на кілька простих та розподілення його виконання між програмними клієнтами. Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованої системи ХПТК ОНПУ дозволило в 8,4 рази скоротити час пошуку гетерогенної інформації.

В результаті впровадження системи пошуку та перетворення гетерогенної інформації в інформаційну систему КТЕП, був скорочений час перетворення гетерогенної інформації на 28%. Час сканування всіх БД та пошук потрібної інформації скоротився на 32%. Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованої системи КТЕП дозволило в 9,1 рази скоротити час пошуку гетерогенної інформації.

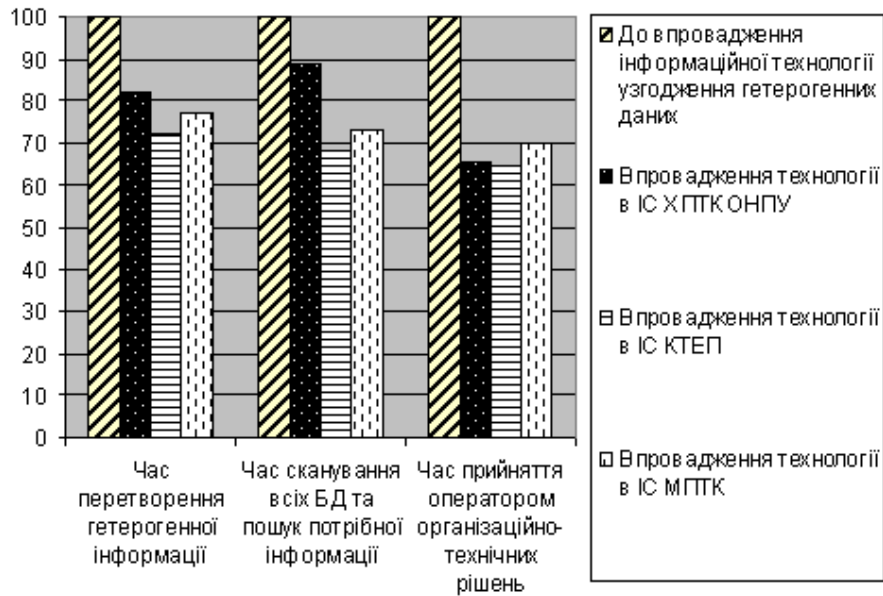


Рисунок 9 – Порівняльна діаграма результатів впровадження інформаційної технології узгодження гетерогенних даних

В результаті впровадження системи пошуку та перетворення гетерогенної інформації в інформаційну систему МПТК, був скорочений час перетворення гетерогенної інформації на 23%. Час сканування всіх БД та пошук потрібної інформації скоротився на 27%. Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованої системи МПТК дозволило в 8,6 разів скоротити час пошуку гетерогенної інформації. На основі моделі оператори, в рамках описаного вище експерименту, витратили на прийняття організаційно-технічних рішень в середньому на 34,8% менше часу в порівнянні з табличним представленням.

Після проведення експериментів в трьох навчальних закладах було встановлено, що середній час перетворення гетерогенної інформації скоротився на 23% (рис. 10).

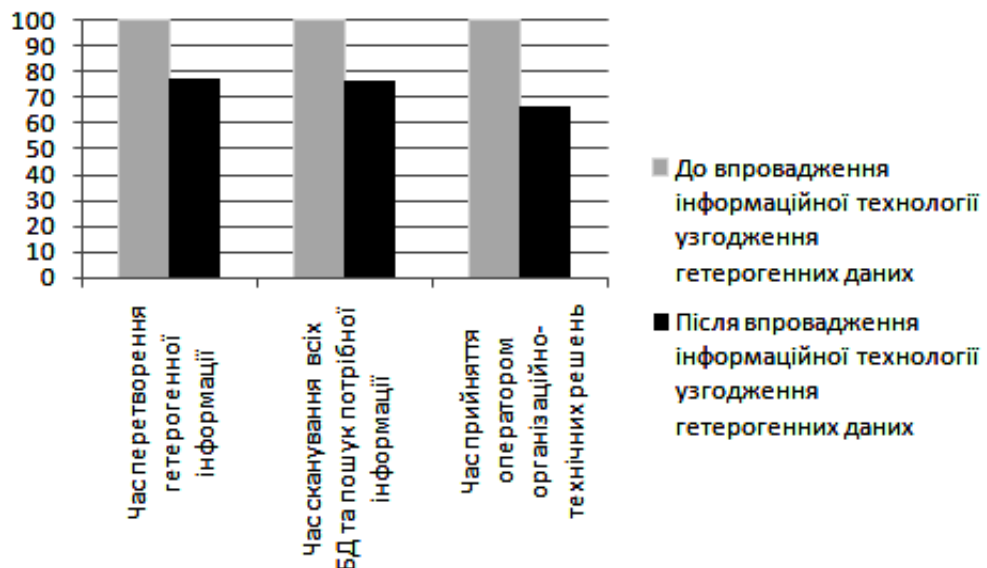


Рисунок 10 – Порівняльна діаграма середнього часу скорочення після впровадження інформаційної технології узгодження гетерогенних даних

Середній час сканування всіх БД та пошук потрібної інформації скоротився на 23,3%. Оператори за допомогою моделі представлення інформації про стан виконання запитів в ІС витратили на прийняття організаційно-технічних рішень в середньому на 33,7% менше часу в порівнянні з табличним представленням даних.

Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованих систем трьох навчальних закладів дозволило в 8,7 разів скоротити час пошуку гетерогенної інформації.

ВИСНОВКИ

В дисертаційній роботі отримано наукові і практичні результати, які в сукупності вирішують проблему дослідження, створення і впровадження системи пошуку та перетворення гетерогенної інформації, як програмного інструмента для організації автоматичного руху потоків даних та виконання складних завдань в гетерогенних розподілених БД.

1. Проведено аналіз існуючих методів перетворення гетерогенних даних в інформаційних системах, проблем передачі пакетів інформації та їх сумісності, тенденцій розвитку сучасних інформаційних технологій. Одним з виявлених недоліків було те, що більшість підрозділів, що мають свою АС, не мають доступу до бази даних інших систем, або дублюють її, що є нераціональним. Було встановлено, що характер даних систем призводить до суттєвого збільшення часу пошуку та перетворення даних, що не відповідає сучасній практиці. Для скорочення часу був обраний програмно-клієнтський підхід.

2. Розроблено рефлексорно-комутаційний метод перетворення гетерогенної інформації, відмінною особливістю якого є використання програмних клієнтів для передачі та узгодження гетерогенних даних, що дозволило скоротити час перетворення. Застосування програмних клієнтів дозволило, використовуючи звичайні локальні БД, виконувати складні запити із сутностями, які розміщені на різних робочих станціях, шляхом поділу складних запитів на кілька простих і відправлення їх відповідним клієнтам.

3. Запропонована тривимірна модель представлення інформації про стан виконання запитів в ІС на основі хромоматематичного підходу, що дозволило скоротити час прийняття оператором організаційно-технічних рішень. На основі моделі оператори, в рамках описаного вище експерименту, витратили на прийняття організаційно-технічних рішень в середньому на 34,8% менше часу в порівнянні з табличним представленням.

4. Вдосконалено метод виявлення взаємозв'язаних списків, що дозволило скоротити час пошуку гетерогенних даних в розподілених інформаційних системах.

5. Розроблено інформаційну технологію узгодження гетерогенних даних з використанням рефлексорно-комутаційного методу перетворення інформації.

6. За допомогою рефлексорно-комутаційного методу перетворення гетерогенної інформації та методу виявлення взаємозв'язаних списків виконано інтеграцію існуючих інформаційних систем ХПТК ОНПУ, КТЕП та МПТК з ЄДЕБО. За допомогою запропонованих методів на процес пошуку та перетворення було витрачено відповідно в 8,4, 9,1 та 8,6 разів менше часу, ніж класичним методом.

7. Дістали подальшого розвитку метод і інформаційна технологія узгодження гетерогенних даних в розподілених інформаційних системах, шляхом використання клієнто-орієнтованого підходу, що дозволило підвищити ймовірність виконання запиту при скороченні часу на його виконання.

8. Розроблена система пошуку та перетворення гетерогенних даних з метою використання рефлекторно-комутаційного методу перетворення гетерогенної інформації в розподілених БД.

9. Проведений порівняльний аналіз роботи інформаційної системи, який показав, що впровадження системи пошуку та перетворення гетерогенної інформації в ХПТК ОНПУ дозволило скоротити час перетворення на 18%. Час сканування всіх БД та пошук потрібної інформації скоротився на 11% за рахунок розділення складного запиту на кілька простих та розподілення його виконання між програмними клієнтами. В результаті впровадження системи пошуку та перетворення гетерогенної інформації в інформаційну систему КТЕП, був скорочений час перетворення гетерогенної інформації на 28%. Час сканування всіх БД та пошук потрібної інформації скоротився на 32%. В результаті впровадження системи пошуку та перетворення гетерогенної інформації в інформаційну систему МПТК, був скорочений час перетворення гетерогенної інформації на 23%. Час сканування всіх БД та пошук потрібної інформації скоротився на 27%.

10. Після проведення експериментів в трьох навчальних закладах було встановлено, що середній час перетворення гетерогенної інформації скоротився на 23%. Середній час сканування всіх БД та пошук потрібної інформації скоротився на 23,3%. Оператори за допомогою моделі представлення інформації про стан виконання запитів в ІС витратили на прийняття організаційно-технічних рішень в середньому на 33,7% менше часу в порівнянні з табличним представленням даних. Впровадження методу виявлення взаємозв'язаних списків для узгодження ЄДЕБО з даними автоматизованих систем трьох навчальних закладів дозволило в 8,7 разів скоротити час пошуку гетерогенної інформації.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Сафонов, М. С. Модель знаходження оптимуму знань-умінь студента в умовах індивідуалізації навчання [Текст] / П. С. Носов, Ю. І. Косенко, М. С. Сафонов // Вісник національного технічного університету / Нові рішення в сучасних технологіях. – №24, Харків: НТУ, 2011 – С. 68-72. [Видання включено до МНБ – eLibrary, Ulrich, Google Scholar, DOAJ]

2. Сафонов, М. С., Моделирование управления потоками данных в информационной системе [Текст] / М. С. Сафонов, А. Е. Яковенко // Сборник научных трудов / Труды Одесского политехнического университета / Компьютерные и информационные сети и системы — Вып. 1 — Одесса, 2013 — С. 97 – 103. ISSN 2076-2429. [Видання включено до МНБ – РІНЦ, Index Copernicus, Ulrich]

3. Сафонов, М. С., Математичне забезпечення методу управління потоками даних [Текст] / М. С. Сафонов, О. Є. Яковенко, В. М. Тонконогий // Збірник наукових праць / Високі технології в машинобудуванні — Вип. 1(23) — Харків: НТУ «ХПІ», 2013 — С. 161 – 165. ISSN 2078-7677. [Видання включено до МНБ – eLibrary, ULRich, Google Scholar, DOAJ]

4. Сафонов, М. С. Метод реінжинірингу інформаційної системи з використанням об'єктів управління [Текст] / М. С. Сафонов // Збірник наукових праць / Електротехнічні та комп'ютерні системи — Вип. 13(89) — Київ, 2014 — С. 105-113. [Видання включено до МНБ – ВНДТІ РАН, Ulrich, eLibrary]

5. Сафонов, М. С. Імітаційне моделювання обробки даних в інформаційній системі [Текст] / М. С. Сафонов // Збірник наукових праць / Східно-європейський журнал передових технологій — Вип. 2/4(68) — Харків: Технологічний Центр, 2014 — С. 37-42. [Видання включено до МНБ – PIHЦ, Copernicus, Ulrich's Periodicals Directory, EBSCO Publishing, Google Scholar, DOAJ]

6. Сафонов, М. С. Об'єктно-орієнтований метод управління потоками інформації в базах даних [Текст] / М. С. Сафонов, О. Є. Яковенко // Збірник наукових праць / Науковий вісник Херсонської державної морської академії — 2014— С. 297-301. [Видання включено до МНБ – Google Scholar, DOAJ]

7. Сафонов, М. С., Аналіз результатів впровадження та тестування системи управління потоками даних в існуючу інформаційну систему [Текст] / М. С. Сафонов, О. Є. Яковенко, // Збірник наукових праць / Сучасні технології в машинобудуванні — Вип. 9 — Харків: НТУ «ХПІ», 2014 — С. 247 – 255. ISSN 2078-7499. [Видання включено до МНБ – eLibrary, ULRich, Google Scholar, DOAJ]

Опубліковані праці апробаційного характеру

8. Сафонов, М. С. Моделирование и оценивание траекторий решения практических заданий на основе дерева принятия решений [Текст] / П. С. Носов, М. С. Сафонов, А. С. Носов// Сборник научных трудов по материалам международной научно-практической конференции / Современные проблемы и пути их решения в науке, транспорте, производстве и образовании . – Том 2. Технические науки. – Одесса: Черноморье, 2006 – С. 13-18.

9. Сафонов, М. С. Закриття доступу до інформації методом вбудовування у додаток за допомогою технології клієнт-сервер [Текст] / М. С. Сафонов, Г. Ф. Сафонова// Сборник научных трудов по материалам международной научно-практической конференции / Научные исследования и их практическое применение. Современное состояние и пути развития . – Том 2. Технические науки. – Одесса: Черноморье, 2010 – С. 66-68.

10. Сафонов, М. С. Тайлова модель створення карти послідовності операцій [Текст] // М. С. Сафонов , В. В Коленко., А. А Рура // I Міжнародна конференція студентів і молодих учених. Сучасні Інформаційні Технології 2011. – Одеса, 2011 – С. 142-143.

11. Сафонов, М. С. Метод параллельной обработки сущностей в базах данных с различной архитектурой [Текст] / М. С. Сафонов // Сборник научных трудов по материалам международной научно-практической конференции / Современные про-

блемы и пути их решения в науке, транспорте, производстве и образовании . – Вып. 4, Том 5. Технические науки. – Одесса: Черноморье, 2011 – С. 55-58.

12. Сафонов, М. С. Модель нахождения оптимума знаний-умений студента в условиях индивидуализации обучения [Текст] / П. С. Носов, М. С. Сафонов, Ю. И. Косенко // Збірник наукових праць за матеріалами Всеукраїнської науково-практичної конференції / Проектна та конструкторсько-технологічна підготовка майбутніх фахівців інженерного та педагогічного напрямів. – Херсон: Айлант, 2011 – С. 117-119.

13. Сафонов, М. С. Оптимізація клієнт-серверного доступу до інформації за допомогою об'єктно-орієнтовного підходу [Текст] / М. С. Сафонов, О. Є. Яковенко // Сборник научных трудов по материалам международной научно-практической конференции / Современные направления теоретических и прикладных исследований. – Вып. 1, Том 4. Технические науки. – Одесса: Куприенко, 2012 – С. 52-55.

14. Сафонов, М. С. Об'єктно-орієнтовний підхід в ступеневій моделі керування структурними елементами [Текст] / М. С. Сафонов, В. В. Коленко, Ю. І. Косенко // Матеріали I Міжнародної науково-практичної інтернет-конференції / Проблеми та перспективи розвитку науки на початку третього тисячоліття у країнах СНД. – Переяслав-Хмельницький, 2012 – С. 234-235.

15. Сафонов, М. С. Об'єктно-орієнтовний підхід при керуванні віддаленими сховищами даних [Текст] // М. С. Сафонов // II Міжнародна конференція студентів і молодих учених. Сучасні Інформаційні Технології 2012. – Одеса, 2012

16. Сафонов, М. С. Об'єктно-орієнтовний підхід при керуванні запитами до декількох станцій [Текст] / М. С. Сафонов, В. В. Коленко // Сборник научных трудов по материалам международной научно-практической конференции / Перспективные инновации в науке, образовании, производстве и транспорте. – Вып. 2, Том 4. Технические науки. – Одесса: Куприенко, 2012 – С. 17-20.

17. Сафонов, М. С. Стійкість недетермінованих стегосистем [Текст] / В. В. Коленко, М. С. Сафонов // Матеріали I Міжнародної науково-практичної інтернет-конференції / Проблеми та перспективи розвитку науки на початку третього тисячоліття у країнах СНД. – Переяслав-Хмельницький, 2012 – С. 232-233.

18. Сафонов, М. С. Метод знаходження неявних зв'язків між атрибутами в різних сутностях баз даних [Текст] / М. С. Сафонов, В. В. Коленко // Сборник научных трудов по материалам международной научно-практической конференции / Научные исследования и их практическое применение. Современное состояние и пути развития. – Вып. 3, Том 5. Технические науки. – Одесса: Куприенко, 2012 – С. 72-75.

19. Сафонов, М. С. Реалізація електронного документообігу [Текст] / В. В. Коленко, М. С. Сафонов, О. В. Нарожний // Матеріали IV Міжнародної науково-практичної інтернет-конференції / Проблеми та перспективи розвитку науки на початку третього тисячоліття у країнах СНД. – Переяслав-Хмельницький, 2012 – С. 196 - 198.

20. Сафонов, М. С., Використання об'єктів керування для оптимізації потоків інформації в мережевих базах даних з різною архітектурою [Текст] / М. С. Сафонов, О. Є. Яковенко, С. О. Савченко // Збірник наукових праць / Інформаційні технології в освіті, науці та виробництві — Вип. 1 — Одеса, 2012 — С. 60-62 .

21. Сафонов, М. С., 3D Обертально-поступальна модель для зв'язаних джерел інформації [Текст] / М. С. Сафонов, О. Є. Яковенко // Збірник наукових праць / Інформаційні технології в освіті, науці та виробництві — Вип. 1(2) — Одеса, 2013 — С. 196-203.

22. Сафонов, М. С., Прогнозування стану показників об'єктно-орієнтованої моделі в інформаційній системі [Текст] / М. С. Сафонов, О. Є. Яковенко // Збірник наукових праць / Інформаційні технології в освіті, науці та виробництві — Вип. 3(4) — Одеса, 2013 — С. 92-98.

23. Сафонов, М. С., Математична теорія стаціонарної черги в об'єктно-орієнтованому методі управління потоками даних [Текст] / М. С. Сафонов, О. Є. Яковенко // Збірник наукових праць / Інформаційні технології в освіті, науці та виробництві — Вип. 3(4) — Одеса, 2013 — С. 114-119.

24. Сафонов, М. С., Розповсюдженні недоліки та проблеми інформатизації в навчальних закладах [Текст] / М. С. Сафонов, В. Д. Яковенко, Г. Ф. Сафонова // Збірник наукових праць / Інформаційні технології в освіті, науці та виробництві — Вип. 4(5) — Одеса, 2013 — С. 192-197.

АНОТАЦІЯ

Сафонов М. С. Методи та технології узгодження гетерогенних даних в розподілених інформаційних системах. – *На правах рукопису.*

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – Інформаційні технології. – Одеський національний політехнічний університет Міністерства освіти і науки України, Одеса, 2015.

Дисертація присвячена побудові інформаційної технології для узгодження гетерогенних даних. Виконано дослідження математичних моделей та методів, які використовуються в автоматизованих системах управління даними. Розроблена математична модель рефлекторно-комутаційного методу перетворення гетерогенних даних в інформаційній системі, яка характеризує часові та швидкісні параметри роботи системи. Розроблений метод побудови інформаційної моделі виконання запитів в ІС на основі хромоматематичного підходу, що дозволило скоротити час прийняття оператором організаційно-технічних рішень. Вдосконалено метод виявлення взаємозв'язаних списків, відмінною особливістю якого є використання рефлекторно-комутаційного методу перетворення гетерогенної інформації, що дозволило скоротити час виконання складних запитів. Розроблена автоматизована інформаційна система, що реалізує метод рефлекторно-комутаційного перетворення гетерогенних даних та формує на основі отриманої інформації модель представлення інформації про стан виконання запитів. Виконано експериментальне підтвердження ефективності застосування створеної автоматизованої інформаційної системи.

Ключові слова: рефлекторно-комутаційний метод, інформаційний потік, управління базами даних, автоматизована система, взаємозв'язані списки, гетерогенні дані.

ABSTRACT

Safonov M. S. Methods and technologies of concordance of heterogeneous information are in the distributed informative systems. - Manuscript.

Dissertation for a candidate of technical science (Ph.D.) degree in specialty 05.13.06 - Information technology. - Odessa national polytechnic university of the Ministry of Education and Science of Ukraine, Odessa, 2015.

The thesis is devoted to creation of information technology for the solution of problems of heterogeneity of data. Proceeding from it, research of mathematical models and methods which are used in automated control systems for data was executed. The mathematical model of a reflex and switching method of management of data flows in information system which characterizes temporary and high-speed parameters of work of system is developed. The method of creation of information model of management of data flows which gave the chance quickly to receive information on a condition of information system in general is developed. The method of search of implicit communications of attributes in different sushchnost with use of a reflex and switching method of management of data flows which increases functionality and mobility of relational databases is developed. The automated information system which realizes a method of a reflex and switching method of management of data flows is developed and forms on the basis of the received information information model of work of system. Experimental confirmation of efficiency of use of the created automated system is executed.

Keywords: reflex-switching method, information stream, management of databases, the automated system, interdependent lists, heterogenous data.

АННОТАЦИЯ

Сафонов М. С. Методы и технологии согласования гетерогенных данных в распределенных информационных системах. – На правах рукописи.

Диссертация на соискание научной степени кандидата технических наук по специальности 05.13.06 – Информационные технологии. – Одесский национальный политехнический университет Министерства образования и науки Украины, Одесса, 2015.

Диссертация посвящена построению информационной технологии для согласования гетерогенных данных. Исходя из этого, было выполнено исследование математических моделей и методов, которые используются в автоматизированных системах управления данными. Разработана математическая модель рефлекторно-коммутационного метода преобразования гетерогенных данных в распределенной информационной системе, которая характеризует временные и скоростные параметры работы системы. Разработан метод построения информационной модели выполнения запросов в ИС на основе хромоматематического подхода, что позволило уменьшить время принятия оператором организационно-технических решений. Усовершенствован метод обнаружения взаимосвязанных списков, отличительной особенностью которого есть использование реффлекторно-коммутационного метода преобразования гетерогенной информации, что позволило уменьшить время выпол-

нения сложных запросов. Разработана автоматизированная информационная система, которая реализует метод рефлекторно-коммутационного метода и формирует на основе полученной информации модель представления информации о выполнении запросов в информационной системе. Выполнено экспериментальное подтверждение эффективности применения созданной автоматизированной системы.

Большинство информационных систем состоят из отдельных автоматизированных систем, которые выполняют определенные задачи и не выходят за пределы своей БД. БД может содержать одинаковую информацию и иметь несколько разных архитектур, несовместимых между собой, что указывает на их гетерогенность.

Объектом исследования является процесс согласования данных в гетерогенной информационной системе.

Предметом исследования являются модели и технологии согласования гетерогенных баз данных в информационной системе.

Проведен анализ существующих методов согласования гетерогенными потоками данных в информационных системах и проблем передачи пакетов информации, тенденций развития современных информационных технологий. Одним из выявленных недостатков было то, что большинство подразделений, которые имеют свою АС, не имеют доступа к базе данных других систем, или дублируют ее, что является нерациональным.

Проведен анализ совместимости гетерогенных данных, который подтвердил необходимость внедрения рефлекторно-коммутационного метода преобразования гетерогенных данных.

Построена информационная технология согласования гетерогенных данных. Проведена интеграция информационной системы с Единой государственной электронной базой по вопросам образования. Разработано алгоритмическое и программное обеспечение компонентов информационной системы. Данное обеспечение внедрено в организационный процесс и экспериментально подтверждено эффективность предложенных методов и технологий.

В результате эксперимента установлено, что внедрение системы поиска и преобразования гетерогенной информации позволило сократить среднее время преобразования на 23%. Время сканирования всех БД и поиск нужной информации сократилось на 23,3% за счет разделения сложного запроса на несколько простых и распределение его выполнения между программными клиентами. Внедрение метода выявления взаимосвязанных списков для согласования СДЕБО с данными автоматизированной системы позволило в 8,4 раза уменьшить время поиска гетерогенной информации. Операторы с помощью модели представления информации о состоянии выполнения запросов в ИС потратили на принятие организационно-технических решений в среднем на 34,8% меньше времени в сравнении с табличным представлением данных.

Ключевые слова: рефлекторно-коммутационный метод, информационный поток, управление базами данных, автоматизированная система, взаимосвязанные списки, гетерогенные данные.