

ТЕОРЕТИКО-МНОЖИННА МОДЕЛЬ ІНФОРМАЦІЙНОЇ СИСТЕМИ В ЗАДАЧІ ЗОВНІШНЬОЇ ОПТИМІЗАЦІЇ SQL-ЗАПИТІВ

УДК 004.657; 004.043

КОСТЕНКО Павло Павлович

аспірант кафедри комп'ютерних та інформаційних систем

Кременчуцького національного університету імені Михайла Остроградського.

Наукові інтереси: оптимізація SQL-запитів, робото-технічні комплекси, алгоритми програмування.

e-mail: ppkostenko@gmail.com

ВСТУП

Незалежно від типу інформаційних систем (ІС), їх загальним призначенням є зберігання та обробка інформації з метою забезпечення інформаційних потреб користувачів. Зручним та високоефективним засобом зберігання інформації в ІС виступають реляційні бази даних (РБД). Системи керування базами даним (СКБД) та відповідно РБД стали невід'ємною частиною сучасних автоматизованих інформаційних систем (АІС). Функціональна розгалуженість, багатомодульність та дискретність в роботі ІС ускладнюють моделювання та математичне відображення процесів, що відбуваються в системі.

АНАЛІЗ ПОПЕРЕДНІХ ДОСЛІДЖЕНЬ Й ПОСТАНОВКА ЗАДАЧІ

Застосування РБД суттєво прискорює надання інформації користувачеві, проте в процесі функціонування ІС спостерігається поступове зменшення ефективності її роботи, зокрема за рахунок збільшення часу обробки запитів користувачів апаратним забезпеченням ІС [1, 2]. Це обумовлено «старінням» логічної та фізичної структури РБД, ускладненням внутрішніх алгоритмів виконання запитів користувачів тощо. При цьому, з огляду на негативні наслідки від затримки надання інформації користувачам системи, важливим є завдання вирішення проблеми підвищення ефективності

роботи ІС, як на етапах їх розробки, так і під час цільового застосування [3]. Серед спектру заходів підвищення ефективності роботи ІС в ході їх активного використання постійного вдосконалення набувають задачі оптимізації SQL-запитів, зокрема зовнішньої оптимізації SQL-запитів до їх виконання засобами СКБД. Відкриттю залишається задача побудови математичної моделі ІС, що надає можливість описати перетворення вхідних запитів до вихідних даних та дозволяє показати процес зовнішньої оптимізації SQL-запиту.

Наразі відомо три підходи до математичного опису ІС: модель інформаційної системи як дискретної динамічної системи з післядією [4, 5], дворівнева модель інформаційної системи як системи масового обслуговування [6], модель інформаційної системи як кортеж моделей її елементів (теоретико-множинна модель) [7, 8].

Модель інформаційної системи як дискретної динамічної системи з післядією більшого вдосконалення набула в роботі [5], де детально досліджувався процес збору й обробки даних, під яким розуміється лінійна чи умовна послідовність подій, що настають за умови деякого формального правила, реалізація якого призводить до досягнення цілі – введення процесу в необхідний стан чи режим.

Для визначення умов найбільш повного і цілісного відображення реальності, Л.А. Краснодубец використовує класичний апарат формалізації складних

систем [9] та представляє інформаційний процес в вигляді деякої дискретної структурованої системи S (1), функціонування якої залежить від її стану в минулому, а час післядії дорівнює T_k .

$$S = (\{X\}, \{R\}, \{Y\}, \{O\}, \{F\}, \delta, r_0, T_k) \quad (1)$$

де $\{X\}_N = \{X^u\} \cup \{X^f\}$ – кінцева множина подій, що мають вплив на процес; $\{X^u\}$ – підмножина подій, пов'язаних з управляючим впливом на процес; $\{X^f\}$ – підмножина подій, не пов'язаних з управляючим впливом на процес; $\{R\}_M$ – множина станів системи (режимів); $\{Y\}_L$ – множина параметрів режиму, приймається за вихід; $\{F\}_M$ – множина функцій виходу; $\{O\}_N$ – множина алгоритмів, пов'язаних з подіями; δ – функція переходів; $r_0 \in \{R\}_M$ – початковий стан (режим); $T_k = [t_{i-k}, t_i]$ – часовий інтервал післядії, k – ціле число.

Використання моделі дискретної динамічної системи з післядією (1) в умовах структурної та функціональної інформаційної невизначеності БД не є можливим, оскільки моделі вимагають детального вивчення об'єкта керування, його функцій переходів, функцій виходів, множини станів.

В роботі [6] зазначається, що відповідно до теорії мереж систем масового обслуговування, ІС перебуває під дією вхідного потоку запитів λ_0 , складається з підсистем S_1, S_2, \dots, S_n . Взаємодія підсистем описується з допомогою матриці передач (2)

$$\Theta = [\theta_{ij}]_{(n+1) \times (n+1)} \quad (2)$$

де θ_{ij} – вірогідність того, що запит після його обробки в підсистемі S_i буде переданий по каналу зв'язку до підсистемі S_j , $i, j = \overline{0, n}$. В якості S_0 виступає зовнішнє середовище, що генерує вхідні запити та отримує результати обробки запитів (дані).

При розгляді ІС на рівні підсистем вважається, що Λ_j – інтенсивність потоку запитів через систему S_j , $j = \overline{0, n}$, $\Lambda_0 = \lambda_0$. Значення $\Lambda_1, \Lambda_2, \dots, \Lambda_n$ при заданій матриці (2) визначається вирішенням системи лінійних алгебраїчних рівнянь виду (3):

$$\Lambda_j = \sum_{i=0}^n \theta_{ij} \Lambda_i, j = \overline{0, n} \quad (3)$$

Кількісними показниками функціонування підсистем розглянутої ІС вважаються: S^* – підсистеми з мінімальною продуктивністю, S^{**} – підсистеми з максимальною продуктивністю при обробці вхідного потоку запитів та середня кількість запитів, що проходять через кожен підсистему $N_i = \frac{\Lambda_i}{\lambda_0}, i = \overline{1, n}$. Підсистемам S^* та S^{**} відповідають інтенсивності потоків запитів Λ^* та Λ^{**} : $\Lambda^* = \min_{1 \leq i \leq n} \Lambda_i$, $\Lambda^{**} = \max_{1 \leq i \leq n} \Lambda_i$, де $\Lambda_1, \Lambda_2, \dots, \Lambda_n$ – розв'язок системи (3).

Модель підсистеми S_i ІС складається з m_i вузлів. Структура кожної підсистеми описується матрицею передач (4).

$$Q^{(i)} = [q_{kr}^{(i)}]_{m_i \times m_i}, i = \overline{1, n} \quad (4)$$

де $q_{kr}^{(i)}$ – вірогідність того, що запит після обробки в вузлі k , буде передано до вузла r ; $\lambda_k^{(i)}$ – інтенсивність вхідного потоку запитів в k -й вузол досліджуваної підсистеми. Інтенсивність потоків запитів $\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_{m_i}^{(i)}$ при заданій матриці передач (4) визначається з розв'язку системи рівнянь виду:

$$\lambda_r^{(i)} = \sum_{k=0}^{m_i} q_{kr}^{(i)} \lambda_k^{(i)}, r = \overline{1, m_i}.$$

Вхідний потік запитів в підсистему S_i визначається як $\lambda_0^{(i)} = \sum_{j=1}^n \Lambda_j \theta_{ji}, i = \overline{1, n}$, де Λ_j – розв'язок системи

(3). Середня кількість запитів, що проходить через кожний вузол підсистеми, визначається як:

$$\alpha_k^{(i)} = \frac{\lambda_k^{(i)}}{\Lambda_i}, k = \overline{1, m_i}, i = \overline{1, n}. \text{ Вважається, що } \tau_k^{(i)} -$$

середні витрати часу на обробку одного запиту в кожному вузлі підсистеми S_i , тоді середній час перебування запиту в підсистемі визначається як:

$$T = \sum_{i=1}^n N_i T_i = \sum_{i=1}^n \sum_{k=1}^{m_i} N_i \alpha_k^{(i)} \tau_k^{(i)}. \text{ Час знаходження}$$

запиту в певному вузлі підсистеми в загальному випадку складається з часу перебування запиту в черзі на виконання та часу виконання запиту: $\tau_k^{(i)} = w_k^{(i)} + u_k^{(i)}$, де $w_k^{(i)}$ – час перебування запиту в черзі; $u_k^{(i)}$ – час виконання запиту.

Адекватність розглянутої дворівневої моделі визначається ступенем відповідності вхідного потоку

запитів Пуассоновському потоку. Для застосування моделі необхідно визначити матрицю передач запиту, що в умовах інформаційної невизначеності структури СКБД та бази даних (БД) практично неможливо, оскільки заздалегідь невідома кількість підсистем, що приймають участь в обробці запиту та механізми їх роботи.

В роботі [7] ІС пропонується розглядати як пару:

$$I = \langle M, Q \rangle \quad (5)$$

де $M = \langle \dots \rangle$ – кортеж, який описує схему відношень БД; $Q = \{Q_k \mid Q_k = \langle \dots \rangle, k = \overline{1, p}\}$ – множина запитів до БД, Q_k – кортеж, що описує структуру запиту та визначає параметри виконання запиту.

Кортеж опису схеми відношень БД може бути представлено базовому вигляді, за схемою Е.Ф. Кодда [10]: $M = \langle R, S \rangle$, де $R = \{R_j, j = \overline{1, m}\}$ – множина відношень БД, m – кількість відношень; $S = \{S_i, i = \overline{1, n}\}$ – множина доменів, n – кількість доменів, $S_i = \{s_{i_1}, s_{i_2}, \dots, s_{i_k}\}$, k – кількість елементів множини. Відношення R_j представляє собою кортеж $R_j = \langle s_{1j}, s_{2j}, \dots, s_{nj} \rangle$, як підмножина декартового добутку елементів множини S , $R \subseteq S_1 \times S_2 \times \dots \times S_n$. Подібна модель є найбільш простою та не відображає внутрішній взаємозв'язок відношень.

Більш детальною є математична модель А.С. Маркова [11], що відображає концептуальну модель даних, математичну структуру даних як сукупність:

$$M = \{D, R, A\} \quad (6)$$

де $D = \{D_i, i = \overline{1, n}\}$ – задана множина (носії структури, домен); $R = \{R_j, j = \overline{1, m}\}$ – відношення, кінцевий набір відношень, в яких знаходяться елементи множини (типова характеристика структури); $A = \{A_i, i = \overline{1, p}\}$ – обмежуючі умови, накладені на відношення (аксіоми структури). Відношення R_j представляє собою кортеж $R_j = \langle d_{1j}, d_{2j}, \dots, d_{nj} \rangle$, як підмножина декартового добутку елементів множини D , $R \subseteq D_1 \times D_2 \times \dots \times D_n$.

В контексті БД елементи даної математичної структури інтерпретують наступним чином. Абстрактній множині D відповідає певна реальна кінцева множина

елементів даних (задана сукупність типів даних). Відношеннями R відображаються взаємозв'язки між реаліями – в формі відношень між записами даних про ці реалії. Самі відношення між записами, в свою чергу, також можуть бути представлені спеціально організованими даними, наприклад реляційними таблицями. Аксіомам A відповідають умови коректності (цілісності, правильності), накладені на БД.

Модель подібна до (6), представлена в роботі [12]. Реляційна схема (7) представляє собою четвірку, що включає: Ω – кінцева множина атрибутів, Δ – кінцева множина доменів, $dom : \Omega \rightarrow \Delta$ – функцію, що асоціює домен з атрибутом, K – множина множин атрибутів з Ω .

$$M = (\Omega, \Delta, dom, K) \quad (7)$$

В роботах А.Б. Кунгурцева [8,13-14] представлено три математичні моделі БД. Модель (8) є вдосконаленою моделлю (6-7) та відображає розподіл атрибутів (полів відношень) на ключові та всі інші. Такий поділ дозволяє виділити елементи, які пов'язують між собою відношення.

$$M = \langle R, P, B, V \rangle \quad (8)$$

де $R = \{R_i, i = \overline{1, k}\}$ – множина відношень БД; $P = \{P_i, i = \overline{1, k}\}$ – множина полів (атрибутів) відношень; $V = \{V_i, i = \overline{1, k}\}$ – множина ключових атрибутів, призначених для зв'язку між відношеннями, автоінкриментні та індексні атрибути; $B = \{B_i, i = \overline{1, k}\}$ – множина не ключових атрибутів.

Модель (8) не дозволяє визначити зовнішні та внутрішні ключові атрибути, крім того така схема не розкриває ієрархічних зв'язків між відношеннями, функціональних залежностей між атрибутами та не відображає множину доменів БД.

Моделі, що частково вирішують зазначені недоліки, запропоновані в роботах: [7] – модель-кортеж, який описує схему відношень БД та містить множини відношень БД, ієрархічних зв'язків між відношеннями, ключових атрибутів, неключових атрибутів, атрибутів відношень, множину зовнішніх ключових атрибутів (9); [8, 14] – модель-кортеж, який додатково враховує функціональні залежності між атрибутами (10).

$$M = \langle R, D, P, B, A, V \rangle \quad (9)$$

де $R = \{R_i, i = \overline{1, n}\}$ представляє множину відношень БД, n – загальна кількість відношень; $D = \{D_k | D_k = (R_1, R_2), k = \overline{1, d}\}$ – множина ієрархічних зв'язків між відношеннями, (R_1, R_2) – упорядкована пара, R_1 – основне відношення, R_2 – підлегле відношення, d – загальна кількість ієрархічних зв'язків; $A = \{A_i, i = \overline{1, n}\}$ – множина атрибутів відношень, $A_i = (a_1, a_2, \dots, a_t)$ – множина атрибутів відношення $R_i \in R$, t – загальна кількість атрибутів відношення R_i ; $P = \{P_i, i = \overline{1, n}\}$, $P_i = (p_1, p_2, \dots, p_p)$, $P_i \subseteq A_i$, $0 \leq p \leq t$ – множина ключових атрибутів; $B = \{B_i, i = \overline{1, n}\}$, $B_i = (b_1, b_2, \dots, b_b)$, $B_i \subseteq A_i$, $0 \leq b \leq t$ – множина неключових атрибутів, $A_i = P_i \cup B_i$; $V = \{V_i, i = \overline{1, n}\}$ – множина зовнішніх ключових атрибутів, $|V_j| \geq 0$; $|V_j| = 0$, якщо не $\exists D_k = (R_i, R_j)$, для відношень R_i та R_j таких, що $\exists D_k = (R_i, R_j)$, виконується рівність $V_j = P_i$.

$$M = \langle R, A, P, V, B, F, D \rangle \quad (10)$$

де шістка $\langle R, A, P, V, B, D \rangle$ чітко відповідає моделі (9); $F = \{F_j, j = \overline{1, f}\}$, $F_j : x \rightarrow y$, $x \in P_i$, $y \in B_i$ – множина функціональних залежностей між атрибутами, f – загальна кількість функціональних залежностей.

Модель SQL-запиту залежить від умов її застосування і може мати акценти на структуру запиту [15], параметри його виконання [16], результат виконання запиту [17], [7], що відображено в моделях (11)-(15).

$$Q = \langle d_{st}, t_{st}, q, t_q, wp \rangle \quad (11)$$

де d_{st} та t_{st} – дата та час початку виконання запиту, q – текст запиту, t_q – тривалість виконання запиту, wp – ідентифікатор робочого місця, що ініціювало виконання запиту. Модель (11) не відображає структуру SQL-запиту і може використовуватися лише для збору статистичної інформації щодо інтенсивності надходження запитів від робочого місця ІС, часу виконання запиту СУБД, тощо.

Модель (12) повно-функціонально відображає структуру запиту та службову інформацію його виконання. Такий підхід представлення, атрибутів та відношень запитів дозволяє відокремити атрибути

результуючого відношення та детально описати ролі атрибутів і відношень в виконанні запиту. Проте недоліком моделі можна вважати параметри виконання запиту, що варто віднести до моделі результату виконання (моделі вихідних даних виконання запиту). Використання моделі (12) можливе лише для статистичної обробки результатів роботи ІС і не можливе в задачі оптимізації SQL-запитів.

$$Q = \langle T_q, W, Typ, \langle t_{st}, \tau \rangle \rangle \quad (12)$$

де $T_q = \langle T_i, \langle F, R \rangle \rangle, i = \overline{1, N_{T_q}}$ – множина таблиць T_i БД, що беруть участь в запиті q , N_{T_q} – кількість таблиць T_i ; $F = \{F_j, j = \overline{1, N_{F_j}}\}$ – множина полів таблиці T_i , N_{F_j} – кількість задіяних в запиті q полів F_j таблиці T_i ; $R = \{R_k, k = \overline{1, N_{R_k}}\}$ – множина ролей поля F_j в запиті q , N_{R_k} – кількість ролей поля F_j ; $R_k \in \{P_R, O_R, G_R, C_R\}$, де P_R – операція виконання в запиті q виведення поля F_j до результуючої таблиці T_i ; O_R – виконання сортування даних за полем F_j ; G_R – виконання групування даних за полем F_j ; C_R – використання поля F_j в умовах відбору рядків таблиці T_i ; W – ідентифікатор робочого місця, з якого направлено запит q до БД; $Typ \in \{TD, TI, TU, TS\}$ – тип запиту q , де TD, TI, TU, TS – запит на видалення, внесення, оновлення, вибір даних відповідно; t_{st} – час запуску запиту; τ – тривалість виконання запиту, запущеного в момент часу t_{st} .

Модель запиту (13) інформативно відповідає моделі (11) та має відповідні недоліки, що унеможлиблює використання даної моделі з задачами синтаксичної оптимізації SQL-запитів.

$$Q = \langle tx, t, G, d \rangle \quad (13)$$

де tx – текст запиту; t – час виконання запиту; G – група, до якої відноситься запит, якщо на основі попереднього аналізу не виявлено приналежності запиту до жодної групи, $G = 0$; d – дата і час початку виконання запиту.

Модель (14) має найкращі характеристики опису структури запиту з усіх розглянутих. Загальним недоліком моделей (11)-(14) можна вважати відсутність кількісних показників синтаксичної структури

ри SQL-запиту, наявність елементів кортежів запиту, що відповідають апостеріорній інформації запиту (час виконання запиту, відносна кількість виконань запиту).

$$Q = \langle s, n_q, T \rangle \quad (14)$$

Кортеж запиту складається із трьох компонентів: s – текст запиту; n_q – відносна кількість виконань запиту; $T = \{Sel, Ins, Del, Upd\}$ – тип запиту (вибірка, вставка, видалення, відновлення). Текст запиту s розглядається як кортеж з п'яти елементів: $s = \langle R^s, D^s, A^s, C^s, f^s \rangle$, де R^s – множина відношень БД, що беруть участь у запиті, $R^s \subseteq R$; D^s – множина ієрархічних зв'язків між відношеннями R^s ; A^s – множина атрибутів відношень із множини R^s , використаних у запиті; C^s – множина умов вибору запиту; f^s – множина виразів, використаних у запиті. A^s , у свою чергу, представляється як кортеж множин $\langle P^s, V^s, B^s, G^s \rangle$, де P^s – множина атрибутів, які використані у запиті й відносяться до первинного ключа; V^s – множина атрибутів, які відносяться до зовнішнього ключа; B^s – множина неключових атрибутів; G^s – множина атрибутів, використаних для угруповання рядків запиту.

Представлення математичної моделі інформаційної системи, як кортежу моделей її елементів дозволяє побудувати гнучку модель, що відобразить фізичну та логічну структуру інформаційної системи та механізму оптимізації запитів. Побудова моделі ІС потребує вдосконалення існуючих математичних моделей, які відображають концептуальну модель БД та структуру вхідного запиту, і вимагає перегляду кількості елементів

кортежу опису ІС з метою максимальної відповідності даної моделі реальній ІС та процесу виконання запитів засобами СУБД.

Теоретико-множинна модель, запропонована в роботах Кунгурцева А.Б. та Зіноватної С.Л., має найкращі характеристики опису структури запиту з усіх розглянутих. Загальним недоліком можна вважати відсутність кількісних показників синтаксичної структури SQL-запиту, елементів кортежів запиту, що відповідають апостеріорній інформації запиту (час виконання запиту, об'єм даних тощо).

Побудова моделі ІС потребує вдосконалення існуючих математичних моделей, які відображають концептуальну модель БД та структуру вхідного запиту, і вимагає перегляду кількості елементів кортежу, опису ІС з метою максимальної відповідності даної моделі реальній ІС та процесу виконання запитів засобами СКБД.

МЕТА РОБОТИ

Розробка теоретико-множинної моделі перетворення вхідних запитів до вихідних даних в інформаційній системі, яка дозволяє описати процес зовнішньої оптимізації SQL-запиту на основі локальної моделі керованого процесу.

ОСНОВНА ЧАСТИНА

Інформаційну систему, що використовує РБД для збереження інформації, пропонується розглядати як систему, що включає три складові: БД, СКБД, робоче місце (РМ) користувача. Відповідно структурно-функціональна схема ІС має вигляд (рис. 1):

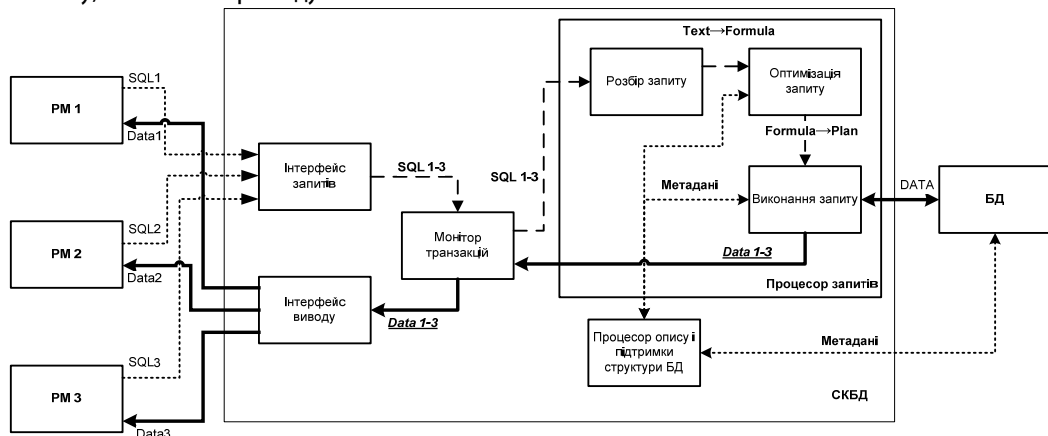


Рисунок 1 – Структурно-функціональна схема ІС

Від користувача системи (PM) надходять SQL-запити, що виконуються в СКБД з урахуванням внутрішніх налаштувань системи та БД. В результаті виконання запиту користувач отримує структуровані дані, відповідно до умов, описаних в SQL-запиті та наявних структур даних в БД. Вважаємо, що в процесі виконання запиту приймають участь наступні складові: $Q = \{Q_i, i = \overline{1, n}\}$ – множина вхідних даних (SQL-запитів), MS – кортеж, що описує об'єкт (СКБД), $Y = \{Y_i, i = \overline{1, n}\}$ – множина вихідних даних (результати виконання SQL-запиту). Тому модель ІС в загальному вигляді представляється кортежем (15):

$$I = \langle Q, MS, Y \rangle \quad (15)$$

Якщо розглядати оптимізацію SQL-запиту в автоматичному режимі до його потрапляння в процесор запитів та без припинення роботи ІС, постає проблема оптимізації запиту в умовах інформаційної невизначеності БД та налаштувань внутрішнього оптимізатора СКБД. За таких умов структурно-функціональна схема частини ІС, що бере участь у обробці SQL-запитів, має вигляд (рис. 2):

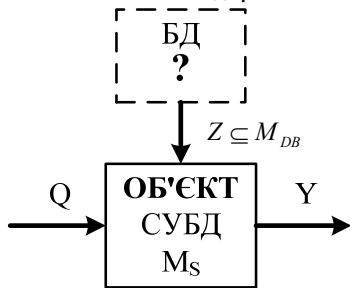


Рисунок 2 – Структурно-функціональна схема частини ІС, що бере участь у обробці SQL-запитів

До моделі ІС вводимо Z – зовнішнє збурення, що діє на об'єкт та представляє собою кортеж, який структурно відповідає моделі БД, кількісно виражає множини даних, необхідних для виконання Q , та характеризує вплив структури БД на виконання запиту. MS містить в собі функції, що узгоджують вхід Q і вихід Y .

$$Y_i = M_S(Q_i, Z_i) \quad (16)$$

де $Q_i = \langle S, T \rangle$ – кортеж, що описує вхідний SQL-запит, $T = \{Sel, Ins, Del, Upd\}$ – тип запиту (вибірка, вставка, видалення, відновлення відповідно); S – текст запиту, представляє собою кортеж семи елементів:

$S = \langle R^S, A^S, Ak^S, A_{RES}^S, C^S, G^S, O^S, F^S, f^S \rangle$, де $R^S = \{R_i^S, i = \overline{1, n}\}$ – впорядкована множина відношень запиту, n – загальна кількість відношень; $A^S = \{A_i^S, i = \overline{1, k}\}$ – впорядкована множина атрибутів запиту, k – кількість атрибутів запиту; $Ak^S = \{Ak_i^S, i = \overline{1, l}\}$ – впорядкована множина комбінованих атрибутів запиту $Ak^S \subseteq A^S, l \leq k, l$ – кількість комбінованих атрибутів запиту; підмножина $A_{RES}^S = \{A_i^S, i = \overline{1, e}\}$ – містить атрибути результуючого відношення запиту, $A_{RES}^S \subseteq A^S, e$ – кількість атрибутів результуючого відношення; $G^S = \{G_i^S, i = \overline{1, l'}\}$ – впорядкована множина атрибутів групування, l' – кількість атрибутів групування, $l' \leq k, G^S \subset A^S$; $O^S = \{O_i^S, i = \overline{1, l''}\}$ – впорядкована множина атрибутів сортування, l'' – кількість атрибутів сортування, $l'' \leq k, O^S \subset A^S$; $F^S = \{F_i^S, i = \overline{1, h}\}$ – множина функцій, застосованих до атрибутів, умови вибору даних, умов сортування та групування даних, h – кількість атрибутів сортування; $F_i^S = \langle FT, \{FP\} \rangle$ – функція представляє собою пару: $FT = \{max, min, count, sum, \dots\}$ – тип функції, $\{FP_j, j = \overline{1, p}\}$ – множина параметрів функції; $C^S = \{C_i^S | C_i^S = \langle \alpha^C, Term, \beta^C \rangle, i = \overline{1, m}\}$ – впорядкована множина умов вибору даних, m – кількість умов. $C_i^S = \langle \alpha^C, Term, \beta^C \rangle \Rightarrow C_i^S = \alpha^C \sim_{Term} \beta^C$ – кортеж умови містить трійку елементів: $\alpha^C = \{A^C, F^C\}$ – множина можливих значень лівої частини умови, яка містить атрибут A^C чи функцію F^C такі, що $A^C \in A, F^C \in F$; $\beta^C = \{A^C, F^C, Q^C, \{\emptyset\}, const\}$ – множина можливих значень правої частини умови, яка може містити атрибут A^C , функцію F^C такі, що $A^C \in A, F^C \in F$; підзапит Q^C , який структурно відповідає $Q_i, Q^C \sim Q_i; \{\emptyset\}$ – не пуста множина значень; $const$ – константне значення; $\alpha^C \subset \beta^C$; $Term = \{=, >, <, \leq, \geq, \neq, in, is, not\{=, >, <, \leq, \geq, \neq, in, is, \}\}$ – множина термів умови; $f^S = \{f_j^S(R_i^S, A_i^S), j = \overline{1, e}\}$ – множина функцій зв'язку відношення запиту R_i^S з атрибутом запиту A_i^S ; $f_j^S(R_i^S, A_i^S): A^S \rightarrow [true, false]$

– функція чіткої належності елемента $A_i^S \in A^S$ до відношення R_i^S .

$M_S = \langle TR, \mu', \mu'', \gamma', \gamma'', MD \rangle$ – кортеж, що описує функціонал СКБД і відповідно містить функції відображення $\mu', \mu'', \gamma', \gamma''$ структури запиту Q_i структурі метаданих БД $MD = \langle R^{MD}, A^{MD}, D^{MD}, ST \rangle$ та моделі БД $M_{DB} = \langle R, D, A, Sh_{DB}, K, K', K'', L \rangle$, де $TR = \{Tr_i, i = \overline{1, n}\}$ – множина активних транзакцій, n – кількість транзакцій; $Tr_i = \langle Q^{Tr}, Y^{Tr}, Info^{Tr} \rangle$ – транзакція складається з трійки, де $Q^{Tr} = \{Q_i^{Tr}, i = \overline{1, m}\}$ – впорядкована множина запитів, які виконуються в межах транзакції Tr_i ; $Y^{Tr} = \{Y_i^{Tr}, i = \overline{1, m}\}$ – впорядкована множина результатів виконання запитів в межах транзакції Tr_i , m – кількість пар $\langle \text{запит}, \text{дані} \rangle$ транзакції Tr_i ; $Info^{Tr}$ – множина службової інформації виконання транзакції Tr_i ; $\mu', \mu'', \gamma', \gamma''$ – функції, що відображають структуру запиту Q_i структурі метаданих БД та моделі БД. Функція μ' – відображення множини відношень запиту Q_i множині відношень метаданих СУБД. $\mu' : R^S \rightarrow R^{MD}$.

Область визначення функції μ' позначимо: $D(\mu') = R^S$, область значень $E(\mu') \subseteq R^{MD}$.
 $\forall R_i^S, \forall R_{i+1}^S = \begin{cases} R_i^S \in D(\mu') \\ R_{i+1}^S \in D(\mu') \end{cases}$, відповідно до $(R_i^S \neq R_{i+1}^S) \Rightarrow \mu'(R_i^S) \neq \mu'(R_{i+1}^S)$
 тотожності сама функція $\mu' : R^S \rightarrow R^{MD}$ ін'єктивна. μ'' – функція відображення множини відношень запиту Q_i множині відношень моделі БД. $\mu'' : R^S \rightarrow R$. γ' – функція відображення множини атрибутів запиту Q_i множині атрибутів метаданих СУБД. $\gamma' : A^S \rightarrow A^{MD}$. γ'' – функція відображення множини атрибутів запиту Q_i множині атрибутів моделі БД. $\gamma'' : A^S \rightarrow A$.
 $R^{MD} = \{R_i^{MD}, i = \overline{1, n}\}$ – множина відношень БД, n – загальна кількість відношень, $R^{MD} = R$;
 $A^{MD} = \{A_i^{MD}, i = \overline{1, k}\}$ – множина атрибутів бази даних, k – кількість атрибутів БД. Множина атрибутів бази даних містить в собі підмножини: P^{MD} – множина атрибутів, які використані у запиті й відносяться до первинного ключа; V^{MD} – множина атрибутів, які

відносяться до зовнішнього ключа; B^{MD} – множина не ключових атрибутів; Idx^{MD} – множина проіндексованих атрибутів;
 $A^{MD} = P^{MD} \cup V^{MD} \cup B^{MD} \cup Idx^{MD}$, $A^{MD} \sim A$;
 $D^{MD} = \{D_i^{MD}, i = \overline{1, m}\}$ – множина доменів бази даних,
 $D_i^{MD} = \{d_1, d_2, \dots, d_\eta\}$ – множина значень домену D_i^{MD} , η – кількість значень домену D_i^{MD} , $D^{MD} = D$;
 $ST = \{St_j | St_j = \langle A_i^{MD}, Val \rangle\}$ – множина статистичної інформації використання даних, описується парою атрибут A_i^{MD} , відносно значення частоти звернень Val .

$Y_i = \langle R^Y, A^Y, t, V \rangle$ – кортеж, що описує результат виконання запиту Q_i , містить: R^Y – вихідне відношення, $R^Y \sim R_i$; $A^Y = \{A_i^Y, i = \overline{1, k}\}$ – впорядкована множина атрибутів вихідного відношення, k – кількість атрибутів вихідного відношення, $A^Y \subseteq A^S$; t – час виконання запиту; V – об'єм даних, які повертаються ініціатору виконання запиту.

$Z = \langle R^Z, D^Z, A^Z, Sh^Z, K^Z, K'^Z, K''^Z, L^Z \rangle, Z \subseteq M_{DB}$ – кортеж, що описує зовнішнє збурення, де $R^Z = \{R_i^Z, i = \overline{1, n}\}$ – впорядкована множина відношень, які приймають участь у виконанні запиту, n – загальна кількість відношень; $D^Z = \{D_i^Z, i = \overline{1, m}\}$ – множина доменів бази даних, які приймають участь у виконанні запиту; $A^Z = \{A_i^Z, i = \overline{1, k}\}$ – впорядкована множина атрибутів БД, які приймають участь у виконанні запиту, k – кількість атрибутів запиту;

Схемою зовнішнього впливу Sh^Z – вважаємо систему схем відношень зовнішнього впливу:

$$Sh^Z = \begin{cases} R_1^Z (A_{11}^Z, A_{12}^Z, \dots, A_{1k_1}^Z) \\ R_2^Z (A_{21}^Z, A_{22}^Z, \dots, A_{2k_2}^Z) \\ \dots \\ R_n^Z (A_{n1}^Z, A_{n2}^Z, \dots, A_{nk_n}^Z) \end{cases} \Leftrightarrow \begin{cases} Sh_{R_1^Z} = \{A_{11}^Z, A_{12}^Z, \dots, A_{1k_1}^Z\} \\ Sh_{R_2^Z} = \{A_{21}^Z, A_{22}^Z, \dots, A_{2k_2}^Z\} \\ \dots \\ Sh_{R_n^Z} = \{A_{n1}^Z, A_{n2}^Z, \dots, A_{nk_n}^Z\} \end{cases} \quad (17)$$

$K^Z = \{K_i^Z, i = \overline{1, n}\}$, $K^Z \subseteq A^Z$ – множина не ключових атрибутів; $K'^Z = \{K_i'^Z, i = \overline{1, n}\}$, $K'^Z \subseteq A^Z$ – множина ключових атрибутів, n – кількість атрибутів, $A^Z = K^Z \cup K'^Z$; $K''^Z = \{K_i''^Z, i = \overline{1, n}\}$ – множина зовнішніх ключових атрибутів,

$K^{nZ} \subseteq K^{iZ} \Rightarrow K^{nZ} \subseteq A^Z,$
 $(\{K^Z\} \cup \{K^{iZ}\} \cup \{K^{nZ}\}) \subseteq Sh^Z;$
 $L^Z = \{L_j^Z | L_j^Z = (R_1^Z, R_2^Z), j = \overline{1, d}\}$ – множина ієрархічних зв'язків між відношеннями, (R_1^Z, R_2^Z) – упорядкована пара, R_1^Z – основне відношення, R_2^Z – підлегле відношення, d – загальна кількість ієрархічних зв'язків. В розглянутій схемі функція відображення μ'' отримує вигляд $\mu'' : R^S \rightarrow R^Z$. Відображення є бієктивним оскільки R^S є еквівалентом R^Z , $R^S \sim R^Z$, $|R^S| = |R^Z|$.

В розгорнутому вигляду модель системи з зовнішнім впливом в якості частини структури бази даних, яка приймає участь в виконанні запиту, має вигляд (17).

$$\begin{aligned}
 \langle R^Y, A^Y, t, V \rangle = & \\
 = \langle TR, \mu', \mu'', \gamma', \gamma'', \langle R^{MD}, A^{MD}, D^{MD}, ST \rangle \rangle & \\
 \left(\langle R^S, A^S, Ak^S, A_{RES}^S, C^S, G^S, O^S, F^S, f^S \rangle, \right. & \\
 \left. \langle R^Z, D^Z, A^Z, Sh^Z, K^Z, K^{iZ}, K^{nZ}, L^Z \rangle \right) & \quad (18)
 \end{aligned}$$

Висновки. Запропонована теоретико-множинна модель перетворення вхідних запитів до вихідних даних в інформаційній системі, яка складається з кортежів, що описують структуру і об'єм даних вихідного відношення, час виконання запиту та містять функції відображення, які описують внутрішні процеси виконання запитів засобами СКБД, дозволяє описати процес зовнішньої оптимізації SQL-запиту на основі локальної моделі керованого процесу.

ЛІТЕРАТУРА:

1. Писарчук О.О. Оцінювання ефективності інформаційних систем за вектором критеріїв /О.О. Писарчук //Збірник наукових праць ЖВІ НАУ. – 2010. – №3 – С.117-123.
2. Методы повышения эффективности внедрения корпоративных информационных систем [Електронний ресурс] /С.А. Алёшкин, компания "Ламинфо" //Технологии корпоративного управления, ITeam. – 2010. Режим доступу: http://www.iteam.ru/publications/it/section_53/article_1869/
3. Мониторинг и повышение эффективности информационных систем: подход Embarcadero [Електронний ресурс] /В. Дудченко //Software-Testing.Ru. – 2008. Режим доступу: <http://software-testing.ru/library/testing/performance-testing/93-embarcadero>
4. Гуляев А.И. Временные ряды в динамических базах данных. – М.: Радио и связь, 1989. – 128 с.: ил.
5. Краснодарец Л.А. К вопросу об управлении процессами обработки данных в глобальных наблюдательных системах /Л.А. Краснодарец //Динам. Системы. – 2000. – №16 – С.187-191.
6. Гинатуллин И.А. Двухуровневая математическая модель корпоративных информационных систем на основе линейных стохастических сетей /И.А. Гинатуллин, В.С. Моисеев //Исследования по информатике. – 2004. – №7. – С.89-100.
7. Зіноватна С.Л. Інформаційна технологія підвищення продуктивності автоматизованих систем методами реструктуризації бази даних: Автореф. дис. на здобуття наук. ступеня канд. тех. наук : спец. 05.13.06 «Інформаційні технології» /С.Л. Зіноватна. – Одеса, 2008. – 18 с.
8. Кунгурцев А.Б. Модель реструктуризации реляционной базы данных путем денормализации схемы отношений /А.Б. Кунгурцев, С.Л. Зиноватная //Тр. Одесск. политехн. ун-та. – Одесса, 2006. – №2 (26). – С.105-111.
9. Бусленко Н.П. Моделирование сложных систем. – М.: Наука, 1968. – 156 с.
10. E.F. Codd. A Relational Model of Data for Large Shared Data Banks. Communications of the ACM, Volume 13, Number 6, June, 1970. – P.379.
11. Марков А.С., Лисовский К.Ю. Базы данных. Введение в теорию и методологи: Учебник. – М.: Финансы и статистика. 2006. – 512 с.
12. Giedrius Slivinskas. Bringing Order to Query Optimization /Giedrius Slivinskas, Christian S. Jensen, Richard T. Snodgrass //ACM SIGMOD Record, Vol. 31, No. 2, June 2002, pp. 5–14. September 2002. A DB Technical Report.
13. Кунгурцев А.Б. Математическая модель объектного представления реляционной базы данных /А.Б. Кунгурцев, А.С. Неизвестный //Тр. Одесск. политехн. ун-та. – 2007. – №1 (27). – С.130-134.
14. Кунгурцев А.Б. Анализ целесообразности реструктуризации базы данных методом введения нисходящей денормализации /А.Б. Кунгурцев, С.Л. Зиноватная //Тр. Одесск. политехн. ун-та. – 2006. – №1 (25). – С.104-109.
15. Кунгурцев А.Б. Имитационная модель таблиц реляционной базы данных /А.Б. Кунгурцев, С.Л. Зиноватная, Мунзер Аль Абдо //Тр. Одесск. политехн. ун-та. – 2011. – №1 (35). – С.90-96.
16. Кунгурцев А.Б. Иерархическая модель объектов для исследования запросов к базе данных /А.Б. Кунгурцев, С.Л. Зиноватная //Тр. Одесск. политехн. ун-та. – 2008. – №2 (30). – С.130-134.
17. Кунгурцев А.Б. Поиск закономерностей в распределении запросов для управления материализованными представлениями /А.Б. Кунгурцев, Ю.Н. Возовиков, Као Ван Ань //Тр. Одесск. политехн. ун-та. – 2008. – №2 (30). – С.135-140.