

ретельне дотримання вимог міжнародних стандартів проведення перевірок та регулярне інспектування підприємств відповідними уповноваженими органами.

Виходячи з вищенаведеного, велика кількість медпрепаратів, регіональність спеціалізованих лабораторій та дороговизну проведення деяких аналізів, можна зробити висновок, що необхідно створювати інформаційну систему для підвищення ефективності роботи лабораторій.

IV. Предмет та об'єкт дослідження

Предметом дослідження є процес в результаті якого відслідковується та контролюється інформація про лікарські препарати починаючи з поступлення препарату на ринок України (реєстрація), та післяреєстраційний нагляд за лікарськими засобами:

а) система контролю за складом лікарських засобів, які поставляються на ринок через Державну інспекцію з контролю лікарських засобів і;

б) відстеженням побічних дій та ефективності ліків через систему фармаконагляду, в межах якої у кожній області працюють представники фармакологічного центру по збору та обробці відповідної інформації.

Об'єктом дослідження в цій роботі є інформаційні потоки, які забезпечують належний контроль лікарських препаратів.

Список використаних джерел

1. Сучасні вимоги до організації роботи лабораторій з аналізу якості лікарських засобів / В.Г. Марченко, С.В. Сур, В.П. Черних та ін. — Х., 2002; Good Practices for National Drug Control Laboratories / WHO Technical Report Series, 2002. — № 902.

УДК 004.6

ПОБУДОВА МАТЕМАТИЧНОЇ МОДЕЛІ ЗАДАЧІ ПРОЕКТУВАННЯ ГІБРИДНИХ СХОВИЩ ДАНИХ З ВРАХУВАННЯМ СТРУКТУР ДЖЕРЕЛ ДАНИХ

Яцишин А.Ю.

НТУУ «Київський політехнічний інститут»

I. Постановка проблеми

При проектуванні сховищ даних постає питання коректної побудови структур даних для ефективного його функціонування.

В різних наукових публікаціях були запропоновані різні математичні моделі для формалізації задачі проектування сховищ даних. Однак модель, яка б враховувала особливості задачі проектування гібридних сховищ з врахуванням структур джерел даних, запропонована не була.

При цьому, актуальним та невирішеним на сьогодні є питання автоматизованого чи напівавтоматизованого проектування гібридних сховищ даних у цілому. Ця проблема впливає з практичної необхідності автоматизованої побудови сховищ даних з оптимальною швидкістю за вибраним критерієм оптимальності сховища.

II. Аналіз останніх досліджень та публікацій

Питаннями з побудови сховищ даних та оптимізації сховищ даних займалися: Wen-Yang Lin, I-Chung Kuo [1], Chuan Zhang, Xin Yao, Jian Yang [2,5], Ladjel Bellatreche, Kamel Boukhalfa [3], J.-T. Horng, Y.-J. Chang, B.-J. Liu [4], Goran Velinov, Danilo Gligoroski, Margita Kon Popovska [6], Michael Lawrence [7], Бакуліна М.А. [8]

У праці [9] розглянуто основні принципи побудови та функціонування сховищ даних. У статтях [10] і [11] розглядаються питання проектування гібридних сховищ даних.

III. Мета роботи

Ціллю роботи є запропонувати математичну модель для задачі автоматизованого проектування гібридних сховищ даних, б враховувала і розміщення даних у сховищі, і оптимізаційні фактори (матеріалізовані подання, індекси, схеми розбиття), частоту виконання запитів до сховища та структури джерел даних.

IV. Виклад основного матеріалу

Для формулювання математичної моделі задачі проектування гібридних сховищ даних з врахуванням структури джерел сховищ даних потрібно зробити формальний опис джерел даних, сховища даних, яке отримуємо в результаті проектування, а також оптимізаційних параметрів та рівнянь оптимізації.

Джерела даних. Джерела даних можуть бути структурованими файлами (розділеними чи XML) та базами даних двох типів: OLAP і OLTP.

У випадку розділених структурованих файлів метадані не задаються. У таких файлах може бути задана лише назви полів даних, що входять у файл. Типи даних не задаються. Формалізуються вони так : $S=\{s_1,s_2,\dots,s_n\}$ – множина атрибутів даних у розподіленому файлі.

У випадку структурованих файлів XML можуть бути задані метадані. У якості метаданих можуть бути задані і метадані: назва поля, тип та інші необхідні характеристики. Формалізуються вони аналогічно до розподілених файлів : $X=\{x_1,x_2,\dots,x_m\}$ – множина атрибутів даних у файлі XML.

У випадку реляційних баз даних структура даних задається відношеннями. Відношення позначаються як $R(r_1,r_2,\dots,r_l)$, де R – відношення, а r_i – атрибути відношення. Якщо два відношення $R_1(r_1,r_2,\dots,r_l,r_{l+1},\dots,r_{l+p})$ та $R_2(r_1,r_{l+1},\dots,r_{l+p},\dots,r_{l+k})$ мають спільні атрибути $r_1,r_{l+1},\dots,r_{l+p}$, то в одному з відношень цей набір атрибутів приймає унікальні значення, а в іншому – є зовнішнім ключем.

У випадку багатовимірних баз даних структура даних задається набором вимірів $D=\{d_1,d_2,\dots,d_q\}$ та мір (можуть формулюватися також як вимір) $M=\{m_1,m_2,\dots,m_u\}$.

Гібридне сховище даних. Характеристики гібридного сховища даних описані в роботі [11].

У математичній моделі гібридне сховище даних представляється наступним чином. Воно містить :

Атрибути даних: $V = \{b_1,b_2,\dots,b_v\} = S \cup X \cup R \cup D \cup M$, тобто атрибути джерел повинні бути присутні в сховищі даних, що проектується.

Таблиці даних, що складаються з атрибутів даних: $T = (\{t_1,t_2,\dots,t_w\} | t_i = \{b_j\}, j \in \overline{1,v})$

Області сховища, що містять таблиці даних:

$$A = \{A=a_1,a_2,\dots,a_o | a_i = \{t_j\}, j \in \overline{1,w}, \exists b_k \in t_{j_1}, b_k \in t_{j_2}\}.$$

Оптимізаційні параметри та рівняння оптимізації.

Цільова функція задачі має вигляд:

$$z = \sum_{i=1}^n t_i(L_a, \{I_c\}, \{M_c\}) + \sum_{j=1}^{n-1} T_j + \left(\frac{1}{F'_a} - \frac{1}{F_a}\right)L_a \quad (1)$$

$$\text{або ж } z = \sum_{i=1}^n t_i(L_a, \{I_c\}, \{M_c\}) + \sum_{j=1}^{n-1} T_j + (T'_a - \hat{T}_a)L_a, \quad (2)$$

де n – кількість таблиць запиту, a – область, що містить таблицю i , $\{I_c\}$ та $\{M_c\}$ – ознаки індексування та матеріалізації полів таблиці i , $F_a = \sum_{i \in a} f_i$ – частота доступу до області a . F'_a – встановлене порогове значення частоти доступу до даних області a , \hat{T}_a – середній час виконання запитів, T'_a – порогове значення часу виконання запиту у до даних області a .

Показник частоти задається тому, що при високій частоті доступу до даних доцільно розміщувати у реляційній базі даних, оскільки бази даних такого типу більш придатні для використання у якості операційних баз даних, тобто підтримці запитів з великою частотою.

Запишемо остаточне формулювання задачі проектування гібридних сховищ даних.

Задані множини атрибутів розділених файлів S , файлів XML X , відношення у реляційній базі даних R , виміри багатовимірної бази даних D та міри M . Крім того, відоме порогове значення частот доступу до даних. Спроекувати гібридне сховище даних, визначивши області сховища даних A , таблиці T та атрибути V .

Знайти такі значення ознак розміщення L_a , індексування I_c , матеріалізації M_c , на яких значення цільової функції (16) буде мінімальним серед всіх можливих наборів значень цих змінних.

Список використаних джерел

1. Wen-Yang Lin. A Genetic Selection Algorithm for OLAP Data Cube, / Wen-Yang Lin, I-Chung Kuo - Knowledge and information systems 2004, vol. 6
2. Chuan Zhang. An Evolutionary Approach to Materialized Views Selection in a Data Warehouse Environment. Systems/ .Chuan Zhang, Xin Yao, Jian Yang - Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on Volume 31, Issue 3, Aug 2001
3. Ladjel Bellatreche. An Evolutionary Approach to Schema Partitioning Selection in a Data Warehouse / Ladjel Bellatreche, Kamel Boukhalfa - Lecture notes in computer science, Congrès DaWaK 2005 : data warehousing and knowledge discovery: International conference on data warehousing and knowledge discovery No7, Copenhagen , DANEMARK 2005 , vol. 3589
4. J.-T. Horng. Applying evolutionary algorithms to materialized view selection in a data warehouse/ J.-T. Horng, Y.-J. Chang, B.-J. Liu - Soft Computing - A Fusion of Foundations, Methodologies and Applications, Volume 7, Number 8 / August, 2003
5. Chuan Zhang. Evolving Materialized views in a Data Warehouse / Chuan Zhang, Xin Yao, Jian Yang - Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on Volume 2
6. Goran Velinov. Framework for Generalization and Improvement of Relational Data / Goran Velinov , Danilo Gligoroski, Margita Kon Popovska - IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.3, March 2008
7. Michael Lawrence. Multiobjective genetic algorithms for materialized view selection in OLAP data/ Michael Lawrence - Proceedings of the 8th annual conference on Genetic and evolutionary computation, 2006
8. Бакулина М.А. Модели и алгоритмы автоматизации проектирования структур хранилищ данных для аналитической обработки числовых показателей
9. Шаховська Н.Б. Сховища та простори даних: Монографія. / Шаховська Н.Б, Пасічник В.В. - Львів: Видавництво Національного університету «Львівська політехніка», 2009. – 244 с.
10. Яцишин А.Ю. Застосування генетичного алгоритму для проектування гібридних сховищ даних [Текст]. Подано у друк до Вісника Національного університету „Львівська політехніка”, секція "Інформаційні системи та мережі"/ Яцишин А.Ю. - м.Львів - 2010
11. Яцишин А.Ю. Підходи та алгоритми проектування гібридних сховищ даних [Текст]. Вісник Національного університету „Львівська політехніка”, секція "Інформаційні системи та мережі" / Яцишин А.Ю - м. Львів - 2010