УДК 004.738.5

А. А. БУРЛАКОВ, Д. С. ЧАРНЕЦЬКИЙ
Хмельницький національний університет

ВИКОРИСТАННЯ МОДЕЛІ ДАНИХ “СУТНІСТЬ — АТРИБУТ — ЗНАЧЕННЯ” ДЛЯ РЕАЛІЗАЦІЇ ГНУЧКОГО ПОШУКУ ІНФОРМАЦІЇ

Проводиться ґрунтовний аналіз існуючих способів організації гнучкого пошуку інформації. Пропонується використання підходу EAV (сутність — атрибут — значення) для підвищення продуктивності пошукових алгоритмів.

There is carried out the fundamental analysis of existing ways for organizing the adaptable search of information. There is suggested the use of EAV approach (essence — attribute — value) for increasing productivity of search algorithms.

Ключові слова: пошук інформації, підхід EAV.

Створення механізму швидкого пошуку інформації за критеріями, що динамічно формуються, є актуальною задачею в різних галузях народного господарства. Особливої значимості це питання набуває в сфері електронної комерції.

Для реалізації поставленої задачі можна виділити два підходи:

1) EAV (сутність — атрибут — значення);

2) FlatTables (звичайний реляційний підхід).

FlatTables – це модель даних, яка представляє собою сукупність двовимірних масивів даних, які складаються зі стовпців і рядків. Модель передбачає додавання нової таблиці у базу даних для кожного нового критерію. Таким чином, можна зробити наступний висновок: використання моделі даних Flat Tables є доцільним тільки у тих випадках, коли критеріїв є відносно небагато, а структура критеріїв відносно постійна і рідко змінюється. Якщо критеріїв багато і вони постійно піддаються модифікаціям, то у даного підходу з’являється велика кількість недоліків, а саме:

1) велика кількість таблиць у базі даних;

2) складність генерування запитів до бази даних, оскільки кількість таблиць може сягати декількох десятків;

3) часте використання команд alter/drop/create значно сповільнює систему;

4) важкість забезпечувати цілісність даних.

Отже, при розумному використанні FlatTables можна отримати великі переваги в продуктивності, але ця модель даних не є ефективною та складною при великих кількостях та частих корегуваннях критеріїв.

EAV являє собою універсальну структуру опису даних, в якій дані зберігаються фактично в одному лінійному списку. У EAV кількість критеріїв, які можуть бути використані, є потенційно великою, а кількість таблиць у базі даних при цьому не змінюється. Розглянемо складові моделі:

1) сутність – абстрактний елемент моделі, який не має чітко виражених характеристик і спроможний додавати до себе будь-які критерії;

2) атрибут – це конкретний критерій (характеристика), що додається до сутності;

3) значення – це елемент, який відображає можливі значення конкретного критерію (характеристики).

Для кращого розуміння підходу наведемо приклад: мобільний телефон (сутність), об’єм пам’яті (атрибут), 1 Gb (значення). З цього можна зробити висновок, що при великих кількостях та частих коригуваннях критеріїв найкраще підходить модель даних EAV, адже незважаючи на потужні запити і велику кількість індексів, ця модель дає дуже велику гнучкість і простоту використання. Але варто зазначити, що EAV – це лише модель даних і продуктивність її залежить від правильності реалізації бізнес-логіки.

Виходячи з проаналізованого видно, що модель FlatTables при збільшенні критеріїв може стати вузьким місцем у системі і на порядок знизити продуктивність. Тому для максимальної гнучкості і можливості розширення в подальшому, оптимальним варіантом для системи пошуку за критеріями є використання моделі EAV. Варто сказати, що один із найбільш потужних інтернет-магазинів “Magento” від компанії “Varien” також використовує модель EAV, але бізнес-логіка “Magento” не є оптимальною, з нашої точки зору, оскільки швидкість роботи “Magento” є досить низькою.

Розглянемо один із запропонованих в роботі варіантів реалізації бізнес-логіки та структури бази даних для інтенсифікації приросту продуктивності. На рис. 1 зображена структура бази даних.

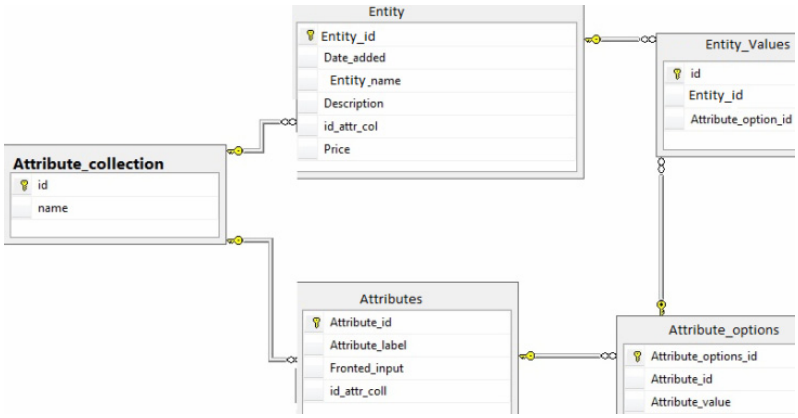

Рис. 1. Схема бази даних для моделі EAV

Ключове значення в схемі даних відіграє таблиця (Attribute_collection), яка розділяє критерії на групи. Сепарація критеріїв на групи дозволяє більш гнучко під'єднувати критерії до сутності, а також зменшує кількість запитів, адже нам одразу відомо, які критерії належать сутності. На рис. 2 наведено блок-схему алгоритму пошуку інформації в базі даних.

Показано лише базову концепцію роботи моделі, однак на цьому модель EAV не вичерпує себе, інкапсулюючи потенціальне розширення. Одним з шляхів подальшого вдосконалення розробки є додання типів даних до критеріїв. Це додає моделі EAV ще більшої гнучкості, адже поряд з типізованими критеріями з'являється можливість пошуку інформації не тільки за точними значеннями, але й в діапазонах.

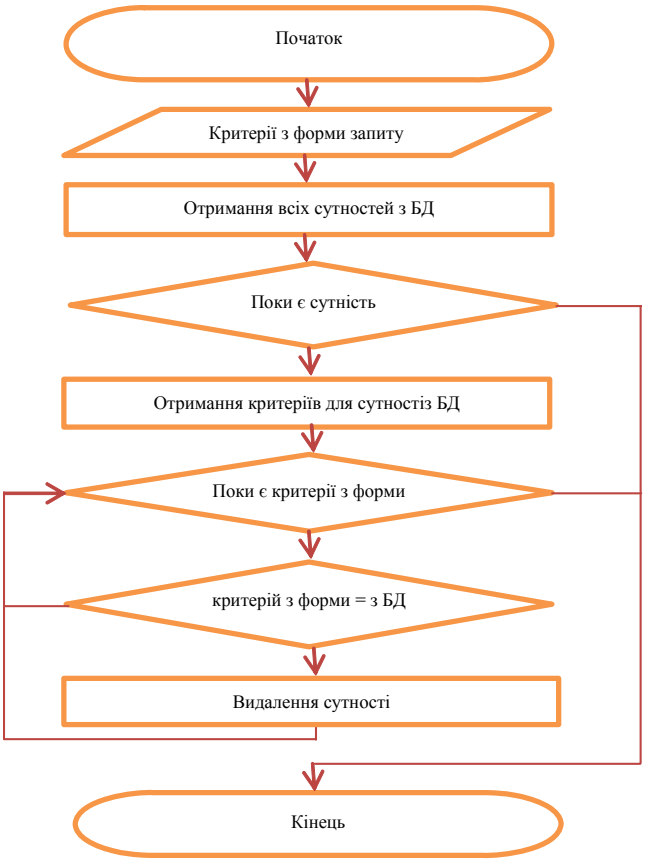

Рис. 2. Пошуковий алгоритм

Аналіз існуючих програмних продуктів показав об'єктивність того факту, що реалізація гнучкого пошуку інформації важлива майже у будь-якому проекті. З розглянутого вище видно, що найбільшої гнучкості пошуку надають алгоритми, які використовують критерії і підтвердженням цієї думки є використання критеріїв у програмних продуктах таких компаній, як "Яндекс", "Varien", "Owox" тощо. Але варто зазначити, що пошук побудований на критеріях не є універсальним засобом, його варто застосовувати тільки у тому випадку, коли у елемента пошуку є чітко виділенні характеристики. Найбільш популярна сфера використання пошуку такого типу – інтернет-магазини та каталоги продукції. З двох варіантів реалізації пошуку, основаного на критеріях,

безперечно кращим варіантом є EAV, оскільки він є набагато гнучкішим за FlatTables і залишає більше можливостей для вдосконалення.

Список використаних джерел

1. Data models [Електронний ресурс]. – Режим доступу : http://knowledge.fhwa.dot.gov/tam/aashto.nsf/All+Documents/4825476B2B5C687285256B1F00544258/$FILE/DIGloss.pdf. – Дата звернення : 28.11.2010.

2. Item Space Data Structures [Електронний ресурс]. – Режим доступу : http://boilerbay.com/infinitydb/ItemSpaceDataStructures.htm. – Дата звернення : 28.11.2010.

3. Magento Database [Електронний ресурс]. – Режим доступу : http://www.magentocommerce.com/wiki/2_magento_concepts_and_architecture/magento_database_diagram. – Дата звернення : 28.11.2010.