

СЦЕНАРІЙ ІНТЕРАКТИВНОГО ПОДАННЯ ЗАПИТУ ДО ГРАФОВОЇ БАЗИ ДАНИХ

Щасливий С.П., Глоба Л.С.

Інститут телекомунікаційних систем НТУУ «КПІ», Україна

E-mail: shchaslyvyi.sergii@gmail.com, lgloba@its.kpi.ua

Scenario for interactive query to the graph database

In the article the problem of intricate for inexperienced users process of querying to the graph databases was considered. A simple scenario for interactive querying to the graph databases, which could simplify this process for above-mentioned users, was presented.

Концепція графових баз даних [1] стає все більш широковживаною і популярною при роботі з різного роду прикладними сценаріями, такими, наприклад, як Семантична павутина (Semantic Web), соціальними і біологічними мережами, а також, базами даних, орієнтованими на роботу з географічними даними. Графова база даних - різновид баз даних з реалізацією мережевої моделі у вигляді графа і його узагальнень[2]. Багато механізмів були запропоновані саме для запиту до графової бази даних, концепція якої, зазвичай, є важкозрозумілою, а інколи, зовсім незрозумілою для недосвідченого користувача, який, не має достатнього знань чи досвіду, щоб задати певний запит до бази даних з використанням спеціальної формальної мови.

Проблема надання допомоги в роботі зі складними базами даних для молододосвідчених користувачів стає все більш актуальною. "Зазвичай, написання певного запиту до бази даних на спеціальній формалізованій мові для користувача займає більше часу, ніж виконання самого запиту, хоч і при цьому не використовуються будь-які внутрішні ресурси бази даних"[3]. У той час як у [3] дана проблема досліджується в контексті реляційних баз даних, можна з впевненістю заявити, що вона стає ще більш актуальною для графових баз даних. Дійсно, графові бази даних, як правило, не мають метаданих, оскільки немає ніякої різниці між екземпляром реалізації і схемою. Також, складною може бути візуалізація, можливо, дуже великих реалізацій.

Для об'ємів даних будь-якого розміру, графові бази даних є кращим способом для представлення і організації запитів до таких даних, що мають певні зв'язки. Зв'язаними є дані, інтерпретація і значення яких вимагає від нас спочатку зрозуміти, яким чином пов'язані їх складові елементи. Найчастіше, для отримання цього розуміння, ми повинні визначити і кваліфікувати зв'язки між речами[4].

Отже, для вирішення проблеми надання допомоги недосвідченим користувачам при роботі з графовими базами даних у правильному формулюванні запиту, потрібно розробити і запропонувати систему, засновану

на сценарії інтерактивного завдання запиту до графової бази даних (див. рис.1).

Наприклад, користувачу в інтерактивному режимі пропонується візуалізувати невеликі фрагменти графа і промаркувати вузли, що представляють інтерес в якості «позитивних» чи «негативних», залежно від того, чи повинні бути ці вузли у складі результуючого запиту. Після кожного взаємодії з користувачем, система відкидає «неінформативні» вузли, тобто ті вузли, які не несуть ніякої корисної інформації для запиту користувача. Таким чином, система також допомагає користувачеві отримати результат запиту з мінімальною кількістю взаємодій. При цьому, важливим є те, що вузол графа вибирається лише в тому випадку, коли до нього існує «шлях», тобто вузли є «сусідами».

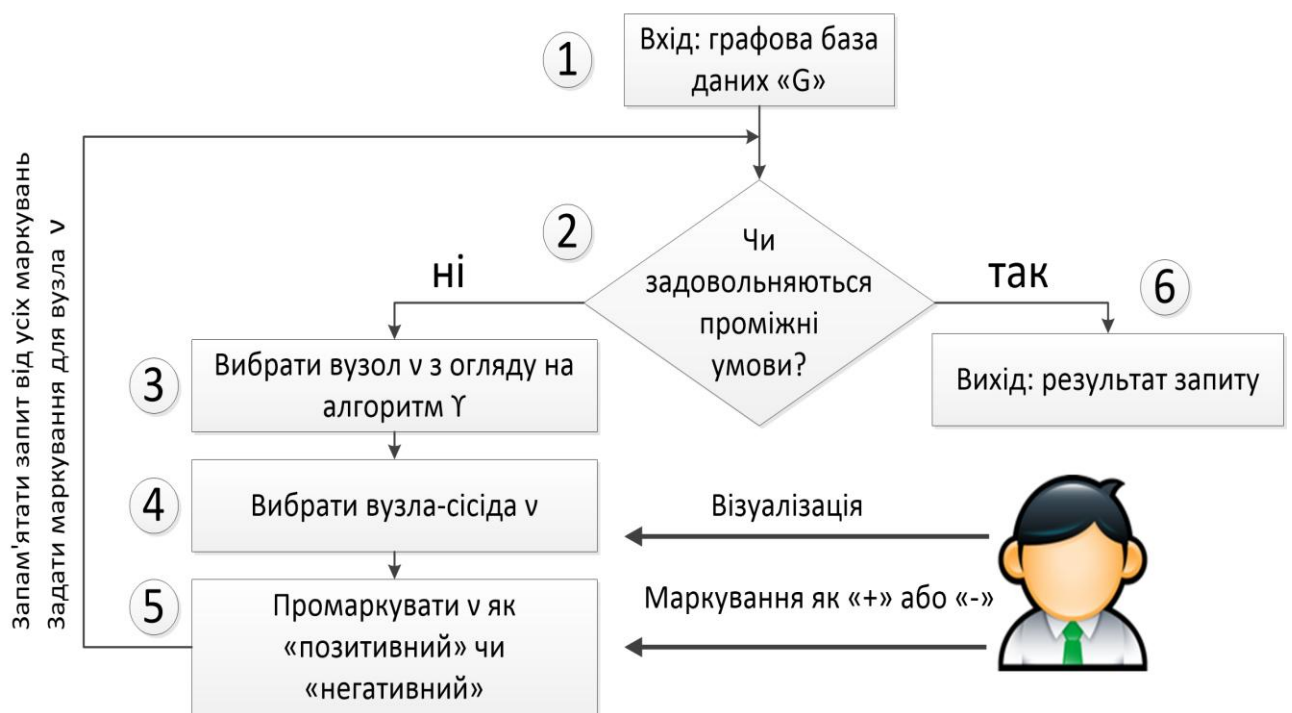


Рис.1 Сценарій шляху проходження запиту графовою базою даних

1-2) Вхідною точкою до умовної графової бази даних є граф G . В початковий момент він уявляється порожнім до початку взаємодії з користувачем. Взаємодія продовжується доти, доки не задоволено заданих користувачем умов. Природним умовою тут є зупинка взаємодії, якщо є тільки один можливий запит, що відповідає умовам для поточного набору прикладів. Користувач також може зупинити процес раніше самостійно, якщо якийсь проміжний етап у процесі взаємодії привів до задоволення умов користувача.

3) Вузли пропонуються до маркування користувачу за певним алгоритмом Y - тобто функцією, яка приймає в якості вхідних даних граф G і безліч прикладів S , і повертає вузол з графа G . Оскільки наша мета полягає в мінімізації зусиль, необхідних для видачі результату запиту користувача,

ефективний алгоритм повинен уникати ситуацій, при яких користувачеві пропонуються ті вузли, які не несуть корисної для результату запиту інформації.

4-6) У графовій базі даних, вузол сам по собі не несе достатньо інформації, щоб дати користувачеві зрозуміти, чи є він частиною результату запиту. Отже, для уявлення повної картини, потрібно знати які вузли з ним з'єднані. Цей крок має на меті отримання невеликого фрагменту вихідного графа, що легко може бути візуалізованим, і, можливо, містить вузли, які користувач зможе промаркувати як «позитивні», чи «негативні». Далі, якщо користувач позначив вузол в якості позитивного прикладу, знаходиться який із шляхів, що включають даний вузол, становить інтерес для системи. З цією метою потрібно побудувати всі шляхи від поточного вузла, які не включають негативно промаркованих вузлів. Після кількох циклів взаємодії користувача і системи, користувачу, в якості результату, видається графове дерево з оптимальним шляхом через нього. При цьому, як проміжний крок, проводиться валідація користувачем отриманого результату на ступінь задоволення висунутим ним до результату умовам.

Отже, при реалізації запропонованого сценарію подання запитів на основі тісної взаємодії системи з користувачем і його безпосередньої участі в формуванні результату запиту до графової бази даних, значно спрощується доступ користувачів до графової бази даних. Користувачі можуть не мати досвіду у написанні запитів адже сценарій не передбачає використання формалізованих мов написання запитів. При цьому немає потреби в візуалізації всього графа, який може бути потенційно дуже великим, достатньо дивитися на ту частину графа, що складається з цікавих користувачу вузлів, а система у процесі взаємодії з користувачем видасть на виході шлях через граф, що задовольнятиме умовам користувача.

Література

1. Query languages for graph databases / P. T. Wood – *SIGMOD Record*, 2012р. – 41(1): ст.50-60.
2. Graph_database [Електронний ресурс] – Електронні текстові дані. – Режим доступу: https://en.wikipedia.org/wiki/Graph_database – 20 березня 2015р..
3. Interactive join query inference with JIM. PVLDB, 7(13) / A. Bonifati, R. Ciucanu, and S. Staworko, 2014 – ст. 1541-1544.
4. Graph Databases /Ian Robinson, Jim Webber, Emil Eifrem – Publisher: O'Reilly MediaFinal Release Date: 2013р. – 224 с.