

УДК 004.651.4

## ЗАСТОСУВАННЯ ФРАКТАЛЬНИХ КРИВИХ ДЛЯ ІНДЕКСАЦІЇ ДВОВИМІРНИХ ОБ'ЄКТІВ

Панасюк О.В.

Науковий керівник – д.т.н., проф. Аврунін О.Г.

Харківський національний університет радіоелектроніки, каф. ШІ  
м. Харків, Україна

тел.: +38(095) 078-56-45, e-mail: oleksandr.panasiuk@nure.ua

The object of research is spatial data structures. The subject of research in the qualification work is the use of fractal curves to traverse space. The purpose of the work is to study the effectiveness of building and accessing data in the structure of spatial data and the impact on the performance indicators of the method of bypassing 2D space.

R-дерево – це деревоподібна структура даних, яка використовується для організації доступу до просторових даних, тобто для індексації багатовимірної інформації, такої, наприклад, як географічні координати, або прямокутники [1, 2]. Типовим запитом з використанням R-дерев може бути такий: «знайти всі таксі в межах кілометра від мене».

За структурою R-дерева дуже схожі на B (або B+) дерева: вузли містять ключі та покажчики на нащадків, кількість ключів в вузлі обмежена (зазвичай, кілька сотен чи тисяч). Для просторових даних ключем є координати обмежуючого прямокутника  $(x_{min}, x_{max}, y_{min}, y_{max})$ . При цьому новий об'єкт має потрапити у той вузол, для якого потрібно найменше розширення його прямокутника.

Таким чином, R-дерево розбиває двовимірний простір на множину ієрархічно вкладених прямокутників, які можуть перетинатись. Основна проблема полягає в тому, що ключі (координати обмежуючих прямокутників) не є лінійно впорядкованими. З цього випливають щонайменше два наслідки: поперше, один й той самий шуканий об'єкт (точка, прямокутник) може належати декільком вузлам. Другим наслідком є те, що ефективність R-дерева дуже суттєво залежить від порядку, за яким воно заповнюється.

Типовою є ситуація, коли вхідні дані є відсортованими по координатам: наприклад, спочатку по  $x$ , а при рівних  $x$  – по  $y$ . В роботі [1] показано, що отримуване «упаковане по  $x$  R-дерево» (рис.1) матиме «полосову» структуру. Якщо враховувати, що довільні вікна запиту є більш-менш збалансованими (тобто квадратними), то така структура дерева призведе до необхідності переглядання великої кількості вузлів, що є неефективним. Для того, щоб мінімізувати кількість перетинів, прямокутники повинні бути якомога близькими між собою за розмірами та якомога квадратнішими. Для цього слід сканувати площину у порядку, який задається фрактальною кривою, яка заповнює простір. Такими є криві Пеано, Гільберта, Z-крива.

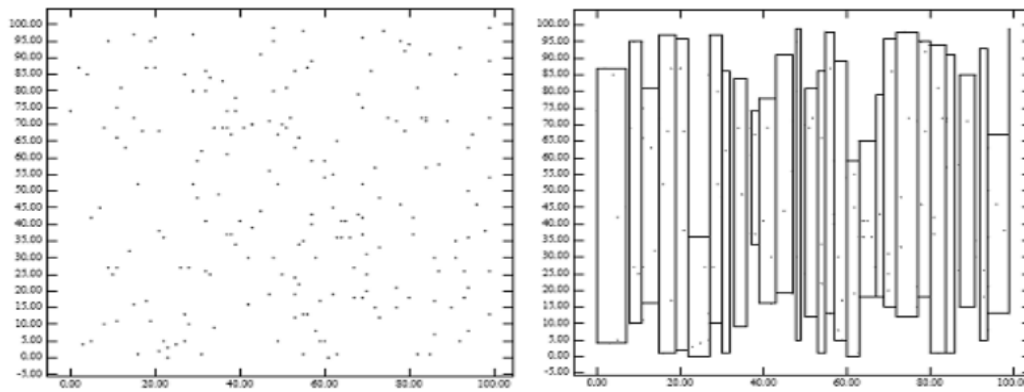


Рисунок 1 – Приклад R-дерева, упакованого по x

Згідно з підходом, що пропонується, кожен об'єкт (точка чи прямокутник) має два ключа: звичайний (координати обмежуючого прямокутника), який використовується під час запитів на пошук даних, та гільбертову відстань (рис. 2). Останній ключ застосовується при модифікації дерев (додаванні та видаленні елементів). В результаті вузли стають більш квадратними незалежно від порядку додавання елементів.

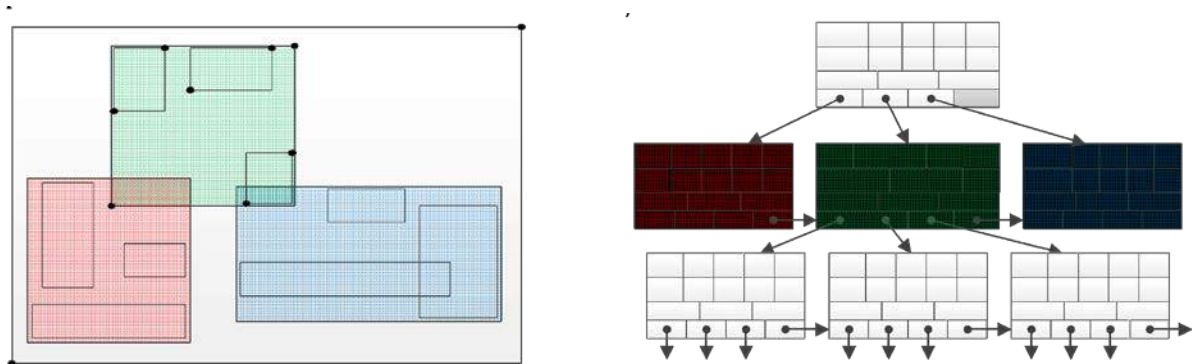


Рисунок 2 – R-дерево з упаковкою за гільбертовою відстанню

Було здійснено програмну реалізацію звичайного та гільбертового R-дерева та досліджено кількість вузлів, що переглядаються, в залежності від розміру вікна запиту. Саме цей показник є головною мірою ефективності, оскільки саме звертання до вузлів є найбільш затратною операцією. Згідно з отриманими результатами, застосування фрактального (гільбертового) обходу площини дає змогу суттєво зменшити кількість вузлів, які проглядаються.

Список використаних джерел:

1. Roussopoulos, N., Leifker D. (1985) Direct spatial search on pictorial databases using Packed R-trees. In Proc. of ACM SIGMOD, pages 17–31, Austin, TX, May 1985.
2. Schubert, E., Zimek, A., Kriegel, H. P. (2013). "Geodesic Distance Queries on R-Trees for Indexing Geographic Data". Advances in Spatial and Temporal Databases. Lecture Notes in Computer Science. Vol. 8098. p. 146.