DS for Range Reporting

1. Computation Model
2. R-Tree
    2.1 Structure

> Yufei Tao. Practical and Theoretical I/O-Efficient Data Structures for Range Reporting. VLDB'15 Summer School Lecture Notes, 2015.

- range reporting query. Let P be a set of N points in $\mathbb{R}^2$. Given an axis-parallel rectangle $q = [x_1, x_2] \times [y_1, y_2]$, a range reporting query reports all the points of P that are covered by q.
    - application: "find all the restaurants in the area";
    - relational database application: TaxPayer(id, sal, age), select id from TaxPayer where 10<=sal<=20 and 50<=age<=60;
- theme: IO-efficient data structure for Range Reporting

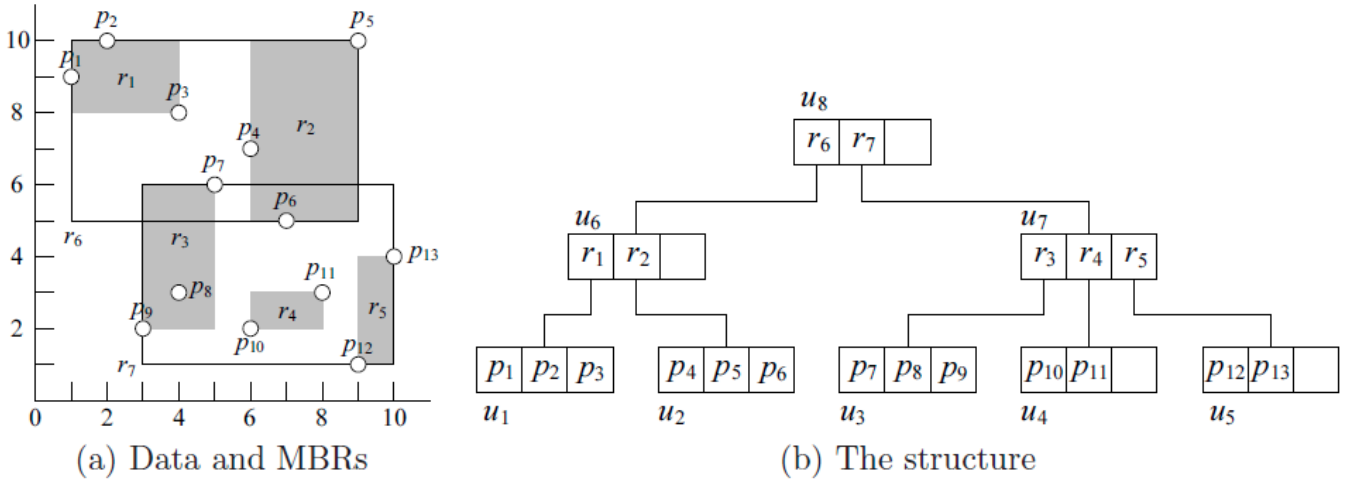
1. Computation Model

- RAM model. An $O(NlogN)$ algorithm means the algorithm is able to solve the problem by performing $O(NlogN)$ "basic operations"
    - standard CUP work or access a memory location
- External memory model. Since an I/O is rather expensive (1-10 milliseconds)
    - space: M words of memory and unbounded disk with block size being B words
    - space complexity: number of disk blocks occupied
    - time complexity: number of I/Os (read B words from disk to memory or write B words conversely), i.e., CPU calculation and memory accesses are free
    - $O(N/B)$ instead of $O(N)$ linear cost

2. R-Tree

> N. Beckmann, H. Kriegel, R. Schneider, and B. Seeger. The R*-tree: An efficient and robust access method for points and rectangles. In SIGMOD, pages 322-331, 1990.

2.1 Structure



(a) Data and MBRs (b) The structure

Figure 1: An R-tree

- Minimal bounding rectangle (MBR): smallest rectangle that tightly encloses all data points, see r_1, r_2, ... in the figure for example
  - good MBR: square-like by reducing perimeter

- leaf node: $b/4$–b points where $b = \Theta(B)$
- non-leaf node: $b/4$–b children, and store an MBR for each child
- $O(N/B)$ space and of $O(\log_B N)$ height