

DSA

Tôn Huỳnh Chí

February 2025

1 Introduction

The R-tree was proposed by Guttman in 1984, and aimed at handling geometrical data, such as points, line segments, surfaces, volumes, and hypervolumes in high-dimensional spaces.

R-trees are hierarchical data structures based on B+ trees (*self-balancing tree data structure used in databases and file systems for efficient indexing and range searches*). They are used for the dynamic organization of a set of d-dimensional geometric objects representing them by the minimum bounding d-dimensional rectangles (*MBRs*). Each node of the R-tree corresponds to the MBR that bounds its children. The leaves of the tree contain pointers to the database objects instead of pointers to children nodes.

It must be noted that the MBRs that surround different nodes may overlap each other. Besides, an MBR can be included (*in the geometrical sense*) in many nodes, but it can be associated to only one of them. This means that a spatial search may visit many nodes before confirming the existence of a given MBR. Also, it is easy to see that

the representation of geometric objects through their MBRs may result in false alarms. To resolve false alarms, the candidate objects must be examined. For instance, Figure 1.1 illustrates the case where two polygons do not intersect each other, but their MBRs do. Therefore, the R-tree plays the role of a filtering mechanism to reduce the costly direct examination of geometric objects.

2 Structure

R-tree is height-balanced tree, similar to B-Tree,