Advanced Database Systems (CS60113)

Assignment 1. R-Tree (Due Date: September 13 2020)

Write C programs to create and query R-Trees. You need to consider only Insert and Search (Delete is not required).

3 separate programs are to be written: One for insert, one for search and one for data creation. The Insert program will save the created R-Tree in a file. The Data Creation program will create and store test data in a file.

Test your code with synthetic data. Generate and Store test data using the Data Creation program in a file (to be used repeatedly). Format will be
d1min d1max d2min d2max d3min d3max … dnmin dnmax
d1min d1max d2min d2max d3min d3max … dnmin dnmax
d1min d1max d2min d2max d3min d3max … dnmin dnmax
..
N such n dimensional rectangles one in each line
Consider each dimension (both min and max) to have only randomly generated integer values between 0 and 20 (for each min, max pair, ensure that max is >= min). Take M=Floor(4096/(4n+1)) and m=Floor(M/2). N=5 Million. Vary n as 2, 5, 10, 50, 100, 200, 500

Overview of steps:
Step 1: Create dataset for n=2. Save it as a file
Step 2: Create R-Tree for the dataset by reading the data stored in the file. Use Insert algorithm with quadratic split for handling node splitting. Save the R-Tree in a file for repeating search.
Step 3: Choose an arbitrary query rectangle in n dimension with each dimension in the integer range 0 to 20 as query rectangle.
Step 4: Implement Search algorithm and query the R-tree (reading it from the file storing the R tree) with the input rectangle. While querying, note the time taken and the number of nodes visited irrespective of whether the rectangle was found or not. Save the data for reporting results.
Step 5: Repeat Steps 3 to 4 fifty times and compute the average time (T) and average number of nodes visited (V). Convert V to the nearest integer.
Step 6: Repeat Steps 1 to 5 for n= 5, 10, 50, 100, 200, 500

Report your results in a table as follows:

Srl. No.	Number of Dimensions (n)	Average Time taken (msec) (T)	Average no. of nodes visited (V)	No. of rectangles (N)	M, m	Linear Search number of nodes N/4096 (L)	L/V
1	2			5,000,000			
2	5			5,000,000			
3	10			5,000,000			
4	50			5,000,000			
5	100			5,000,000			
6	200			5,000,000			
7	500			5,000,000			

Submit your three C programs and a pdf file containing the result in the above format.