CS618: Assignment 4

Total Marks: 100

Due on: 27th February, 2015, 11:30pm

This assignment is to help understand the basics of multi-dimensional data-partitioning indexing using *R-trees*.

Implement a d-dimensional *R-tree* or *R*-tree*. The space is $[0, 1]^d$.

Enable it to handle point queries, range queries, kNN queries and window queries. Ensure that it can support insertions as well. (Deletions may be ignored.)

The configuration for the R-tree must be read from `rtree.config`.
The first line contains a single value which is the page size in *bytes*. The second line mentions the dimensionality of the space.

You can assume the underflow parameter to be half the overflow parameter.
Use the file `assgn4_r_data.txt` to inject the points. It contains 10^6 2-dimensional points.

Use the file `assgn4_r_querysample.txt` to read the queries. The queries have the following formats:

Operation	Code	Details	
Insertion	0	Point	
Point query	1	Query point	
Range query	2	Query center	Range
kNN query	3	Query center	Number of nearest neighbors
Window query	4	Bottom-left point of query box	Top-right point of query box

Ensure that the implementation is truly disk-based and *not* simulated.

Enable the program to output timing results string from the reading of a query to solving it. Do *not* include the time to print it.

Report the following times for both the structures and for each type of operation: (i) minimum, (ii) maximum, (iii) average, (iv) standard deviation.

What do you conclude?

Submit the program and the answers through the submission portal only. You must name your submission `studentno_assgn4.zip`. The student numbers (which are *not* the roll numbers) are 2-digit codes and are available from the course website.

We will evaluate the program by running a query file with the same format as the sample one. Marks will be deducted for wrong answers.