

COMP3210/6210: Assignment 2

Semester 1, 2022

School of Computing
Macquarie University

Marks: 100 marks (20%)

Due Date: 5:00 PM, 03 June 2022 (Friday, End of Week 13)

What to Submit: Source code, a recorded video, and a report

Where to Submit: Electronic submission via iLearn

This is an individual coding assignment. The objective is to implement the R-tree. Each submission will be graded based on correctness. The rest of the document explains the details.

How Your Submission Will Be Tested:

[Dataset]: You will be given a *dataset* that contains 2D points. The dataset will be provided in a text file as the following format:

```
id_1 x_1 y_1  
id_2 x_2 y_2  
...  
id_n x_n y_n
```

Every line gives a point's id, x-, and y-coordinates. Your program should build an R-tree in memory from the dataset.

[Range Query]: You will be given a set of 100 range queries in a text file whose format is:

```
x_1 x'_1 y_1 y'_1  
x_2 x'_2 y_2 y'_2  
...  
x_100 x'_100 y_100 y'_100
```

That is, each line specifies a query whose rectangle is $[x, x'] \times [y, y']$. Then, we will measure its query efficiency as follows.

You should output to a disk file:

- Firstly, your program should display the *time* of answering queries by reading the entire dataset sequentially. This time serves as the *sequential-scan benchmark* to be compared with the cost of your query algorithms that leverage the R-tree.
- Secondly, display the *number* of points returned by each query-note: we need only *the number of points retrieved*, instead of the details of those points.
- Thirdly, display the *total running time* of answering all the 100 queries, and the *average time* of each query (i.e., divide the total running time by 100).

[Programming Language]: Python, Java, C++ (including variants like C, C#, ...), or any other language approved by the instructor. You can implement the R-tree by using the existing libraries provided in the programming language of your choice (i.e., some standard libraries or the libraries

for R-Tree).

[**Deliverables**]: Your submission includes the following components:

1. **Source Code:** The code you have developed yourself. Make sure your code *can be run* in the standard general programming environment.
2. **Report:** Your report should include the following:
 - A brief description of the main functions in your source code.
 - A clear specification of the requirements for executing your code such as, OS environment, placement of input files, any input parameters, etc.
 - A detailed analysis for the construction and search of an R-Tree.
3. **A Recorded Video:** You need to prepare for a recorded video to show the process of running your program, the delivered results, and a brief introduction of how did you design your program.
4. Zip all *your source code* and the *recorded video* into *a single file*, and name the file in the following format: *yourstudentid_surname.zip*. You need to submit *your zip file* and *your report separately* via the submission system.

Marking: Your total mark earned for this assignment is based on:

- [The Program: 60 marks]
 - **Correctness: 40 marks.**
 - * [Sequential-Scan Based Method (10 marks)]: You need to provide detailed comments for your source code (2 marks). Then, if your program correctly answers m (out of 100) queries by reading the entire dataset (reading all the data points) sequentially, you get $8 \cdot (m/100)$ marks for this part.
 - * [R-Tree Based Method (30 marks)]: You need to provide detailed comments for your source code (5 marks). Then, if your program correctly answers m (out of 100) queries by searching the R-Tree, you get $25 \cdot (m/100)$ marks for this part.
 - **Efficiency: 10 marks.** If the average query time is at least *5 times* faster than the sequential scan, you get 10 marks for this part. If at least *2 times* faster (but less than 5 times), you get 5 marks. If less than 2 times faster, no marks.
 - **Recorded Video: 10 marks.** The video should clearly show the process of running your program (2 marks), show the delivered results (2 marks), and clearly introduce your idea in the design of your program to show your understanding of implementing the R-Tree (6 marks).
- [The Report: 40 marks]
 - **Function Description: 5 marks.** If your report includes a clear description of all the functions in your source code, you get 5 marks. If only part of your functions is introduced, you will be given the marks based on the proportion of the correct answers.
 - **Requirement Description: 5 marks.** If your report includes a clear description of the requirements for executing your code such as, OS environment, placement of input files, any input parameters, etc, and your report includes the *screenshots* of the running results (e.g., the average execution time of both sequential-scan and R-Tree based methods, etc.), you get 5 marks.
 - **Analysing the Working of R-Tree: 30 marks.** Select no less than 10 data points from the given dataset, and one query from the given queries. You need to draw figures to illustrate the whole process of R-Tree construction (20 marks) and the R-Tree based

search (10 marks). The search should traverse several nodes of the tree, and during the construction of the R-Tree, there should be an overflow and a node splitting.

Late Submission: Late submissions will be accepted but will incur a penalty unless there is an approved Special Consideration request. A 12-hour grace period will be given after which the following deductions will be applied to the awarded assessment mark: 12 to 24 hours late = 10% deduction; for each day thereafter, an additional 10% per day or part thereof will be applied until five days beyond the due date. After this time, a mark of zero (0) will be given. For example, an assessment worth 20% is due 5 pm on 1 January. Student A submits the assessment at 1 pm, 3 January. The assessment received a mark of 15/20. A 20% deduction is then applied to the mark of 15, resulting in the loss of three (3) marks. Student A is then awarded a final mark of 12/20.

Zero Tolerance for Cheating: You are required to develop the program and write the report by yourself. If you use some libraries and/or contents from other sources, you need to cite and provide the corresponding references. ***All submissions will be checked for plagiarism.*** Any confirmed cases will be reported to the faculty for disciplinary actions, and the marks of the assignment and the unit will be held until the case has been resolved. See Macquarie University's code of conduct for the details of the process. <https://students.mq.edu.au/study/getting-started/student-conduct>