**Data Science - Exercises**
Institut für Informatik, Christian-Albrechts-Universität zu Kiel
WiSe 2023/24
Prof. Dr. Matthias Renz, Christian B. Niko P., Yannick W.

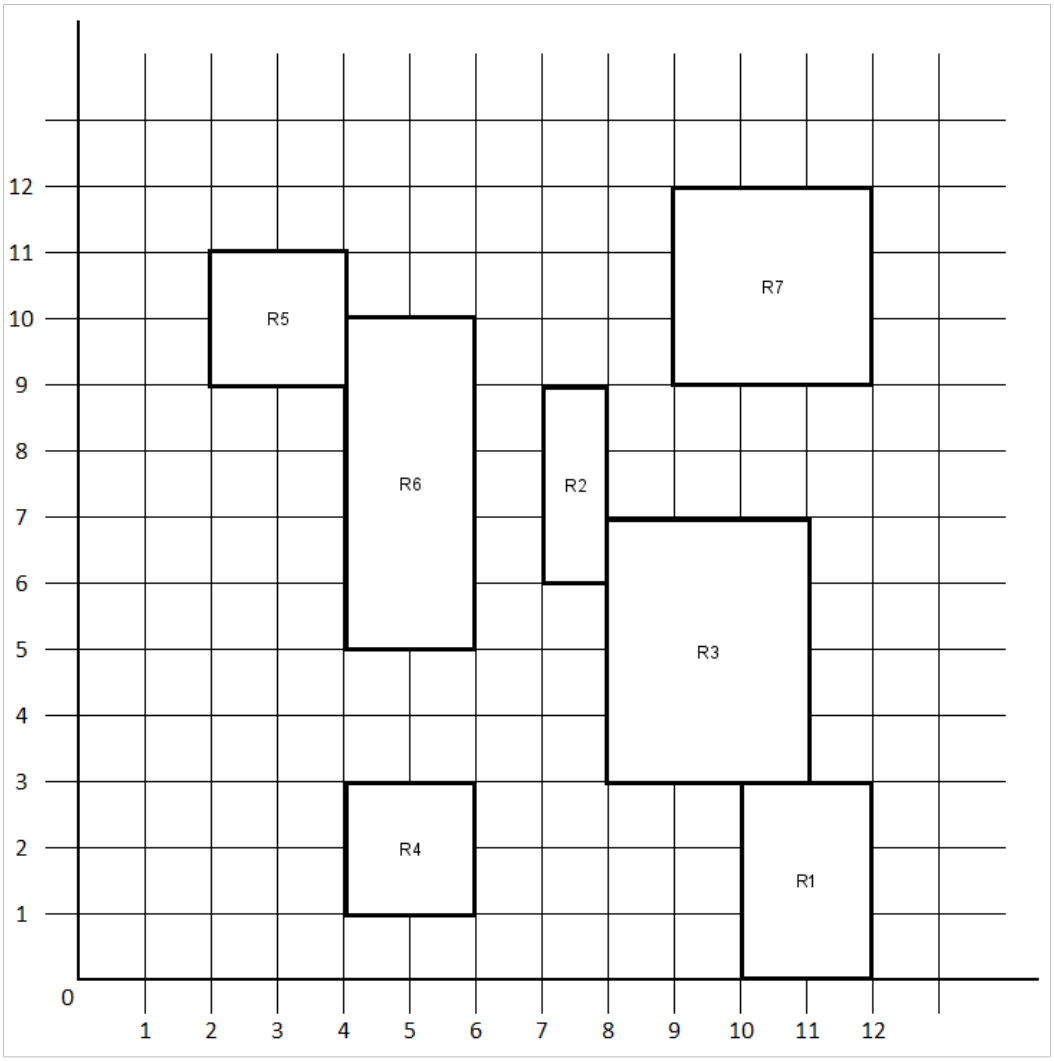
# Data Science Assignment 4 - Search

## 0 Info

- Please hand in one pdf covering all the tasks.
- Only one upload per Team, please.
- You need to hand in sensible solutions for at least 2 of the 3 tasks in order to "pass" this assignment.
- Deadline: 2024-January-31, 23:59am

# 1 R-Tree

Given are polygons that are approximated by the following MBRs R1 to R7:



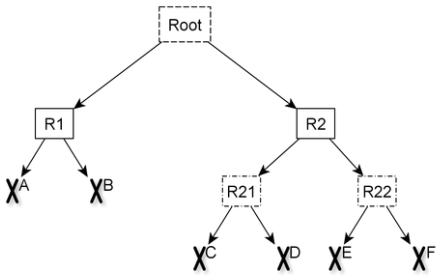

Insert the rectangles in the order R1 to R7 into an R-tree (min = 2, M = 3). After each insert operation, give the coordinates of the newly created or changed page regions as well as the changes to the R tree (as a tree structure). You may give the region coordinates, by drawing the current state of the R-tree after each insert operation onto the given graph (recommended).

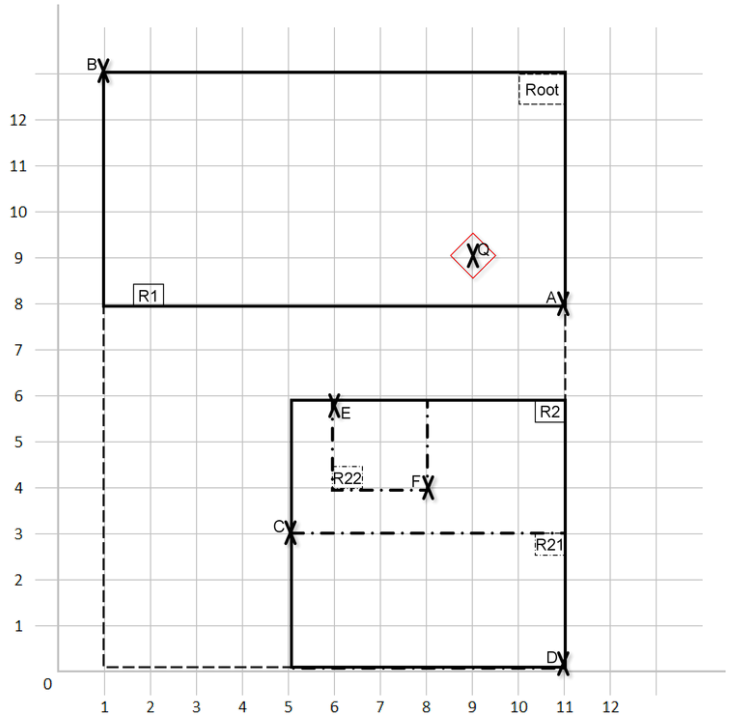
**Insertion strategy**: Use the strategy of the R-tree from the lecture (minimal increase in area) when inserting.
**Split strategy**: When splitting, use the (quadratic) algorithm from the lecture. Please state the two seed regions that were chosen at each split.

Always choose the elements in ascending order (R1,R2,...), when the algorithms do not define which one to choose next.

# 2 $k$NN-Index-APL

Given the objects (A,B,..,F) organized in the following R-Tree and a query object Q:



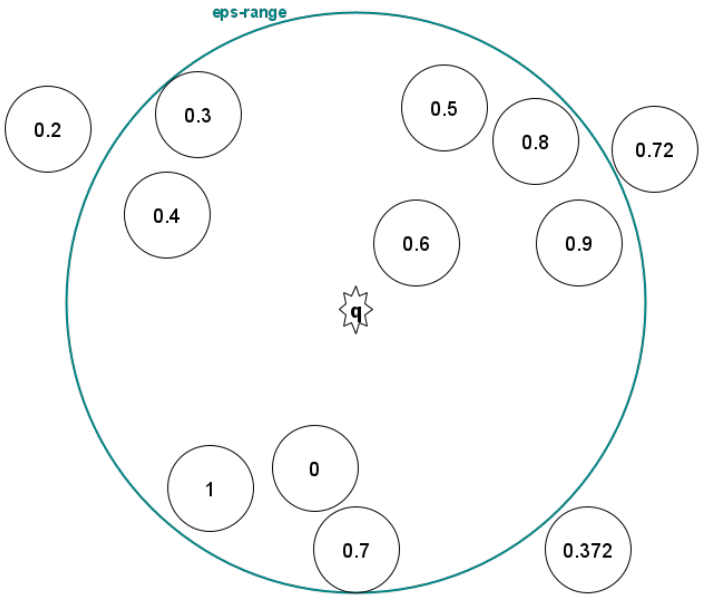

1. Apply the $k$NN-Index-APL algorithm with $k = 3$ and query object Q.
   After each iteration (of the outer while loop), give the current state of the *apl* and the *result* list

2. Reviewing the previous task, please show explicitly where and how we saved database accesses in comparison to a sequential scan, or explain why we didn't.

Use the L1 norm.

# 3 Probabilistic Count Queries

Given the following set of objects, each one marked with its probability of existence, we want to determine the number of objects within the marked epsilon range around query object q:




1. If we wanted to enumerate all relevant possible worlds, how many possible worlds would we have? Please state a number and give a brief reasoning.
2. Apply the polynomial method from the lecture to compute a probability distribution for the number of items in the given range around q.