Boston University

Computer Science Department
CS562 – Advanced Database Applications
Midterm Exam
October 28, 2014

Return *BOTH* question-sheet and answer-booklet

Instructions

Write your name *AND* Student Number on BOTH question-sheet and answer-booklet.

- Write Your Name Here: _____
- Write Your Student Number Here : _____

Spatial Databases - Problem 1 [Points: 30]

1. Consider the point $P = (2, 5.5)$. Give the Z-value of P if we use $K = 4$ bits per dimension. The address space is the square $[0, 16]^2$ (i.e. each dimension ranges from 0 to 16.) Also, give the Hilbert value of P for $K = 2$ bits per dimension. (You can give only the binary values.)

2. Consider again the data space $[0, 10]^2$ and a rectangle with a low-left corner $(6, 0.5)$ and upper-right corner $(9, 5.5)$. Compute the Z-values of the rectangle using $K = 3$ bits per dimension.

3. Compute the two-dimensional pixel (cell) that corresponds to the Z-value 44 when the precision per dimension is $K = 3$. Compute the two-dimensional pixel (cell) that corresponds to the Hilbert value 11, when the precision is $K = 2$.

4. Describe briefly two similarities and two differences between the Grid File and the KD-tree.

5. What are the advantages and disadvantages of the R-tree compared to the Linear Quadtree for indexing regions in a two dimensional space.

Spatial Databases - Problem 2 [Points: 30]

1. Consider the following dataset S of 16 objects in a 2-dimensional space: a=(1,3), b=(1,4), c=(2,0), d=(1,7), e=(2,5), f=(2,8), g=(3,4.5), h=(3,1), i=(4,2), j=(4,3), k=(5,1), l=(5,3), m=(6,1), n=(6,7), o=(7,0), p=(7,1). Also, assume that you have an R-tree built on dataset S, with the following nodes: The root node contains R6 and R7, R6 contains R1, R2, R3 and R7 contains R4 and R5. The MBRs for each node from R1-R7 are (each MBR is defined using each lower and higher point of the main diagonal): R1= [(2,0), (3,1)], R2= [(5,0), (7,1)], R3 = [(4,2), (5,3)], R4 = [(1,3), (3,5)], R5 = [(1,7), (6,8)], R6 = [(2,0), (7,3)], R7 = [(1,3), (6,8)]. The dataset is shown in figure 1.

Give the sequence of pages searched and the results for the following queries:

- Nearest Neighbor query for query point $Q_1 = (3, 2.5)$.
- Window range query with MBR, $Q_2 = [(3.5, 2), (5, 4)]$.

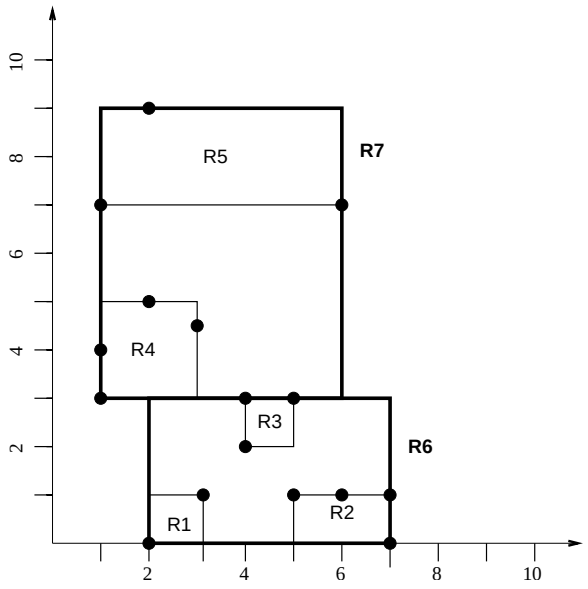


Figure 1: The dataset S and the MBRs of the R-tree.

2. Compute the answer to a Skyline query for this dataset and explain how to find the answer to this query using the R-tree. Show which nodes of the R-tree you need to visit in order to answer the query.

3. Briefly explain how the plane sweep algorithm is used for efficient spatial join processing using R-trees.

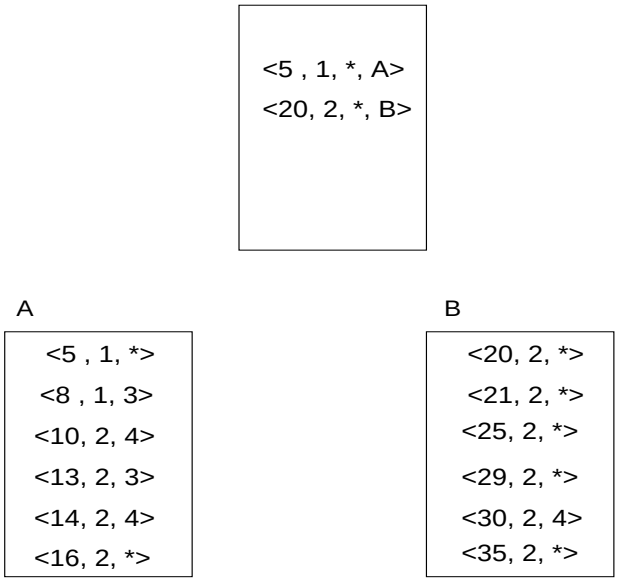
Temporal Databases - Problem 3 [Points: 20]

1. Consider the following temporal evolution of a dataset S:

Time	Oid	Oper
1	a	ins
2	b	ins
4	c	ins
8	d	ins
9	c	del
10	e	ins
12	d	del
15	b	del
20	f	ins
21	e	del
23	g	ins
25	g	del
27	h	ins
28	f	del
30	i	ins

Give the Access Forest for the main memory Snapshot Index, after the execution of the last operation. Show how to answer a timeslice (time-instant) query at the time $t_q = 24$ using the access forest. (Report all the nodes in the structure that you have to visit.)

2. Consider the following instance of a MVBT-tree. We assume that $b = 6, d = 2, k = 3, \epsilon = 0.5$ and the current version is version 4.

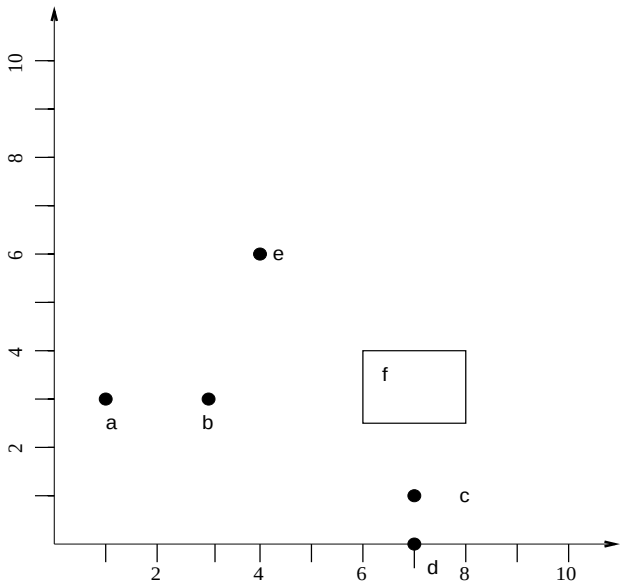


Assuming that the next operation is **Delete 16**, give the structural changes and the new state of the MVBT-tree.

Spatio-temporal Databases - Problem 4 [Points: 20]

1. What are the advantages and dis-advantages of the dual transformation for indexing moving objects?

2. Consider the following dataset S, that has 5 points: a=(1,3), b=(3,3), c=(7,1), d=(7,0), e=(4,6) and one rectangle f=[(6, 2.5), (8, 4)].



Give the Time Parametrized answer and the influence time (T_{INF}) of each object for the following moving queries:

- NN query starting at point PQ = (4,2) and moving with velocity (vx, vy)=(1, 0).
- Window range query starting at the window QW = [(3,1), (5,3)] and moving with velocity (vx, vy)=(1, 0).