This test is a time-limited exam. Please be brief, precise and honest. Good luck!

Please be aware that I will be subtracting points for incorrect answers given in Part 1 and 2 (no negative points in Part 3!).

Note: Make sure that the whole statement is True before you mark it as T. There is no such thing as "partially true statements", they are simply false...

**PART 1. True/False Questions – +/- 6% of the total test’s score**

**Note:** Everything, what does not look like T or F to me, is going to be classified as “– 1 pt.”

- **F** (1 pts.) Abbreviation “SDBMS” used in the context of this course solely stands for “Specially Disturbing Behavior of My Superiors”.

- **F** (1 pts.) ER Spatial Diagrams and UML Class Diagrams are primarily utilized to reflect continuous and amorphous spatial phenomena, e.g. wetlands, snowfalls, clouds, etc.

- **F** (1 pts.) Topological relationships change each time we use different coordinate system. For instance, two skyscrapers which touch each other in a planar paper map will not continue to do so in spherical globe maps.

- **T** (1 pts.) A function Buffer() called on a geometric object returns a geometric region centered at the object. The size of this region is determined by a parameter, which is passed when the Buffer() is called.

- **F** (1 pts.) One way to avoid wastage of disk storage when we utilize Grid Files organization is to apply uniform grid to divide the space, where our data is located. Cells of uniform size guarantee that disc sectors will be always filled almost evenly.

- **F** (1 pts.) In regular R-trees all interior nodes are rectangles. These rectangles are allowed to overlap. This is the reason that no data objects need to be duplicated across the tree’s leaves. Since no data objects are duplicated, each time when a Point Query is performed only one node’s child needs to be followed down the tree.

---

CS 535, Fall 2009 – Test 1
PART 2. Single Choice Questions – +/- 10% of the total test’s score

Note: Everything, what I have hard time identify as a letter, is going to be classified as “− 1 pt.” You need to choose one answer only!!!

1. (3 pts.) Assume you are administrating the MSU’s Registrar Database System and have a large, rarely modified relation of students in the MSU’s database. It is stored in the form of ordered file where SSN is used for data ordering. SSN is a primary key for this relation. You would like to speed up retrieval of data when SELECTs, based on the student’s banner number (i.e. the unique number you have printed on your student ID card), are performed. What type of indexing would you apply?

- a) □ Primary index with block pointers
- b) □ Clustering index with block pointers
- c) □ Secondary index with block pointers only
- d) [X] Secondary index with blocks of record pointers

Assuming every banner ID is unique.

2. (1 pts.) Mark the operation that is topological:

- a) [X] Union,
- b) □ Area
- c) □ Distance
- d) □ North-West

3. (3 pts.) Which statement is true about an R*Tree:

- a) [X] The tree is always balanced, but data objects’ duplicates may appear in the leaves of the tree.
- b) □ The tree is always balanced, and data objects’ duplicates never appear in the leaves of the tree.
- c) □ The tree does not have to be always balanced, and data objects’ duplicates may appear in the leaves of the tree.
- d) □ The tree does not have to be always balanced, but data objects’ duplicates never appear in the leaves of the tree.

4. (3 pts.) Which statement is true about an R+Tree:

- a) [X] The tree is always balanced, and the MBRs which are generated for internal nodes and placed at the same level of the tree, are always disjoint.
- b) □ The tree is always balanced, but the MBRs, which are generated for internal nodes and placed at the same level of the tree, may be overlapping.
- c) □ The tree does not have to be always balanced, but the MBRs which are generated for internal nodes and placed at the same level of the tree, are always disjoint.
- d) □ The tree does not have to be always balanced, and the MBRs, which are generated for internal nodes and placed at the same level of the tree, may be overlapping.
PART 3. Practical Problems – 84% of the total test’s score

1. (14 pts.) Fill out the 1st space below with Z-curve, and the 2nd space with Hilbert curve. Enter values of spatial indexes the curves generated into each of the cells. Shadowed areas reflect Range Queries. Translate shape of the Range Query into (1) a set of z-values, and (2) a set of h-values.

(a) Range Query in z-indexing = \(\left(12, 13, 14, 15\right)\)

(b) Range Query in h-indexing = \(\left(8, 9, 10, 11\right)\)

2. (12 pts.) Country Vatican, denoted below as \(V\), is completely surrounded by another country – Italy, denoted by \(I\). Assume that both countries are modeled by polygons in 2-dimensional space and fill out dimensions of intersections generated by this spatial dependency (i.e. DE-9IM). You may write down your assumptions if you think it is necessary.

<table>
<thead>
<tr>
<th>(V^0)</th>
<th>(\partial I)</th>
<th>(I^-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>(-1)</td>
<td>(2)</td>
</tr>
<tr>
<td>(-1)</td>
<td>(1)</td>
<td>(-1)</td>
</tr>
<tr>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
</tbody>
</table>

Assuming Vatican is a sovereign country where Italy has no influence.
3. (12 pts.) Write down a Spatial SQL query that prints out the names of all rivers which overlap with the Mississippi river (make sure to omit the Mississippi-Mississippi pair). Here is an example of the River data table, you are expected to use:

<table>
<thead>
<tr>
<th>RIVER</th>
<th>Name</th>
<th>Length</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rio Grande</td>
<td>3000</td>
<td></td>
<td>LineStringID-1</td>
</tr>
<tr>
<td>Mississippi</td>
<td>6000</td>
<td></td>
<td>LineStringID-2</td>
</tr>
<tr>
<td>Rio Parana</td>
<td>2600</td>
<td></td>
<td>LineStringID-3</td>
</tr>
<tr>
<td>St. Lawrence</td>
<td>1200</td>
<td></td>
<td>LineStringID-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

```sql
SELECT R2.Name FROM River R1, River R2
WHERE Overlap(R1.Shape, R2.Shape) = 1
AND R1.Name != R2.Name
AND R1.Name = 'Mississippi'
```

By the way, I like to prefer 'Unlike' when we work with text.
4. (18 pts.) Based on the presented below system requirements create ER diagram with spatial pictograms for small Real Estate Business. Make sure to use Pictograms (where necessary), mark primary keys, and write down all necessary Min and Max cardinality constrains:

1. Each house is identified by its unique Address (which is a composition of 4 fields: Line1, City, State, ZipCode). The house has two alternative spatial properties: (1) a location – a centroid point (used for preliminary data search, and display on maps with a large scale), and (2) a polygon (reflecting the actual shape of the house and its exact placement within the lot, when close-up view is generated).

2. Every house has an owner(s), which is identified by his/her SSN (there can be up to N owners of the same house, but there has to be at least one to put it for sale on the real estate market). Additional description of the owner consists of FirstName, LastName, and a PagerNumber.

3. Every house in the Real Estate Database has its (one and only one) ListingAgent, who is identified by an AgentID. FirstName, LastName, and PhoneNumber of the agent need to be stored as well. The location of the agent (e.g. his/her office, home, etc.) needs also to be stored (represented as a point). The agent may have from 1 to N locations, as they are representing different places, he/she can be found in.

4. Exact Date and Time of each listing must be maintained in the system.

5. Every house is located on a Lot, which has its unique LotID. Shape of the lot needs to be stored; it is represented as a polygon. Single Lot can have multiple houses, but we do not store information about lots that have no homes. The house can belong to one and only one Lot. Buyers sometimes request information about Lot’s area, so this information needs to be available.

Write down all additional requirements (if you needed them to complete your diagram). Make sure however that they do not collide with the ones specified above. **Please, make sure your diagram is readable (redraw on the back if needed)**.
Assuming that I am using inverted notation.

Lot ID

Lot

1:N

Located on

Date

Time

TIME_DIM

Lot ID

Acre

Lot

1:1

Address

Additional assumptions not specified

(2)

Listing Agent

Agent ID

First Name

Last Name

Phone Number

Owner

ZIP

State

MUNICIPAL

OWNER

Contact No.

First Name

Last Name

Address

(2)
5. (14 pts.) The picture below presents 7 MBRs of the spatial data objects to be inserted to a spatial database. Following the schemas below explain what changes are going to happen when we use Grid File the data organization. Write pointers names in the Grid Directory, draw grid lines and add scale values (only!) when necessary, and make sure to update the data pages as the objects are inserted. Assume that maximum disk page/bucket capacity = 4 objects. The first split should be created using a HORIZONTAL line.

Here is the situation right after Object 4 was inserted:

Please, start your work from here - After Object 5 is inserted, we have:

After Object 6 is inserted, we have:

After Object 7 is inserted, we have:
6. (14 pts.) We talked in the class about possible different orderings of quads (i.e. nodes) in the Quad Trees. The most popular is ring-like organization (SE, NE, NW, and finally SW), but we also analyzed Z-curve based ordering.

Draw a Region Quad Tree for all 16 objects presented below, using Hilbert-curve type of ordering of quarters in the tree nodes. Be careful – with H-curve the ordering of subquarters may change depending on their location/level. You are expected to use letters representing the quarters when drawing the tree (write them on the branches and make sure to order them properly), and list the data objects in the proper nodes. Make sure to store all objects properly. Assume maximum leaf capacity = 4 objects.