International Institute of Information Technology, Hyderabad

Spatial Data Science (GS2.503)
Spring 2025

Time: 90 minutes **Total Marks: 45**

Note: If needed, make Suitable assumptions and state them clearly when answering. No doubts will be clarified during the exams. Use figures appropriately to better communicate your idea and approach.

Part I. MSQs **10 x 2m= 20m**

[Choose the correct answer/s. One or more answers may be correct]

1) Select properties that are topologically invariant
   a. IIIT campus area is around 66 acres
   b. CR Rao road is connected to the entry gate
   c. Manhattan distance of Vindhya canteen from the entry gate
   d. New construction site inside the campus
2) Select all valid components of a geographic reference system based on latitude and longitude.
   a. Spheroid + Datum
   b. Equator
   c. Prime meridian
   d. Valid projection system
3) By digitizing we convert the geographic features into their digital representation such that they can be stored and processed by a computer. What are the topological constraints you wouldn't apply while digitizing agriculture farms in a village?
   a. Two neighbouring farms can overlap
   b. Two neighbouring farms cannot overlap
   c. A boundary line segment of the farm can have more than two neighbouring farms (including itself)
   d. Corners of the farm may belong to itself and other neighbouring farms
4) Which statements are not true regarding using ray-tracing algorithm for a "point in a polygon" problem?
   a. The time complexity of the algorithm is O(n log n)
   b. If the polygon is convex then the time complexity will grow to $O(n^2)$
   c. The polygon segment must be tested for the collinearity with the half-line originating from the query point
   d. Minimum bounding rectangle can be used to reduce the run time of the algorithm.

5) Which are the core components of a SDBMS?
   a. Spatial Abstract Data Types, i.e. Spatial ADTs
   b. Spatial query language and indexing
   c. Capability of terrain analysis (e.g. slope/aspect) and flow analysis (e.g. shortest path)
   d. Availability of built-in data types such as int, varchar, etc.

6) Classify the following operations into local, focal, and zonal (Hint: draw a figure for better representation)
   a. Derived "class" map from an input temperature map, where each grid cell of a class map is labelled as low, medium and high-temperature location.
   b. Correcting (filling) missing grid-cell values in the elevation raster map using neighbouring values

7) For a transportation network map such as metro network, select the statements that are correct (Metro map here refers to the map displayed within a metro train with info regarding the preceding and succeeding stations)
   a. Metro map is topologically equivalent to the actual network on the ground
   b. Metro map correctly represents the distance between two stations
   c. Connectedness between two stations is often misrepresented in the metro map
   d. A metro line on the ground with a curved shape like the letter "S" can be represented as a straight line in the map

8) For a topological space S and a subset X of S, the near-points of X are the points whose ___ neighborhood contains ___ points of X. Select correct option to complete the previous sentence
   a. (all, at the most one)
   b. (some, some)
   c. (every, some)
   d. No of the above

9) Select statements that are valid for a Field-based model of spatial data
   a. Field-based model better describes the discreet geographic objects
   b. Focal operation uses grid cell values from a 3*3 neighborhood of a grid position
   c. A raster map with the "bigger value" of the spatial resolution provides more details
   d. Temperature, precipitation, and elevation can be better represented by field model

10) In an ER model adapted for representing spatially the districts within a state, which of the following relationships can be true
   a. State boundary is composed of districts
   b. District and State have only a non-spatial relationship like administrative hierarchy
   c. District is contained within a State boundary
   d. District area is comparable to State area

Part II. Answer the following **4 x 3m = 12m**

11) Draw and name the spatial configurations corresponding to the followin

12) g 9IM models. (I: interior, E: Exterior, B: Boundary, T: True, F: False, *: either true or false doesn't matter)

	I	B	E
I	T	*	T
B	*	*	*
E	*	*	*

(A)

	I	B	E
I	T	*	F
B	*	*	F
E	*	*	*

(B)

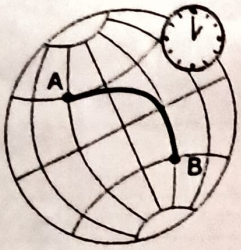
	I	B	E
I	F	F	*
B	F	F	*
E	*	*	*

(C)

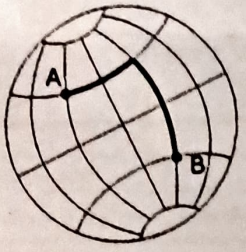
12) About metric spaces.

a) Define when a point set X is qualified as a Metric space.

b) For following examples, state whether they are metric or non-metric space and justify.



travel time distance



Manhattan distance

13) Provide an example application for a zonal operation

a. Describes an input dataset, specifying what each grid cell represents. If you are using another input field data describe that too.

b. Describe the zonal operation and how to derive the output map.

14) For handling spatial data, why do we need a Field model? Argue whether the object-based model is sufficient or insufficient in the following scenarios:

a. to represent the footprint of the houses

b. to represent population density at each location in an urban area marked as residential, industrial, commercial zones.

Mid Semester Exam

25th Feb 2025

Part III. Answer any two of the following 2 x 5m = 10m

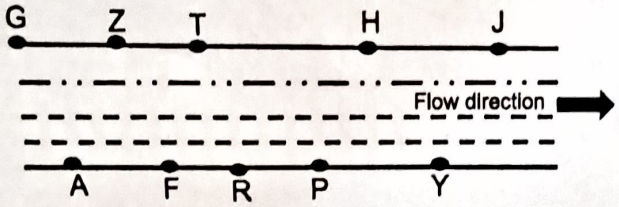
15) A new capital city, similar to the new capital of Andhra Pradesh, Amaravati, was planned adjacent to a major river and forest area to provide a good environment and access to natural resources. The city has expanded over the years and the current city municipal region is an area extending on both sides of the river. [Hint: representation with a figure can help]

a. Write the 4I intersection matrix to identify the "a Highway within the city crossing the River" topological relation

b. Identify a topological condition which requires the 9I intersection matrix, given the above use case. Mention it clearly and construct the matrix for the same.

16) Four NGOs have undertaken a volunteering initiative to clean sections of the Yamuna River in phases. In the first phase, a specific stretch of the river has been selected for cleaning. Based on volunteer's availability, four teams started their cleaning activities on different days. Each team began from a specific location and proceeded **rightward** along the riverbank. At the end of the week, each team reported their **start and end locations**, which are provided in the table below. The task is to identify locations or regions along the riverbank that received **less attention** during the cleaning process.

- Suggest an appropriate **data structure** to efficiently store and analyze the cleaning coverage. (*Hint: showing the data structure will be useful*)
- Provide a **specific query example** to determine an under-cleaned region and outline the **steps** to answer the query.



Team	Start location	End location
1	G	T
2	Z	J
1	T	J
3	A	R
2	F	Y
4	F	P

17) A telecom company has location data (latitude, longitude) for all existing cellular network towers. Due to an increasing number of customer complaints regarding poor network coverage, the company has decided to install new towers to improve coverage. To make an informed decision, the company needs to identify regions with poor coverage and determine the optimal locations for new towers.

a. From "spatial" perspective, what factors contribute to poor coverage?
b. Provide a step-by-step approach to identify underserved areas?
c. Identify the data structures and algorithms that you proposed to use and Comment on the computational complexity of the solution/algorithm?