

Contents

List of Figures	xi
List of Tables	xv
Preface	xvii
Foreword	xxi
Foreword	xxiii
1 Introduction to Spatial Databases	1
1.1 Overview	1
1.2 Who Can Benefit from Spatial Data Management?	2
1.3 GIS and SDBMS	3
1.4 Three Classes of Users for Spatial Databases	4
1.5 An Example of an SDBMS Application	5
1.6 A Stroll through Spatial Databases	11
1.6.1 Space Taxonomy and Data Models	11
1.6.2 Query Language	12
1.6.3 Query Processing	13
1.6.4 File Organization and Indices	16
1.6.5 Query Optimization	18
1.6.6 Data Mining	19
1.7 Summary	20
Bibliographic Notes	20
2 Spatial Concepts and Data Models	22
2.1 Models of Spatial Information	23
2.1.1 Field-Based Model	24
2.1.2 Object-Based Models	26
2.1.3 Spatial Data Types	26
2.1.4 Operations on Spatial Objects	27
2.1.5 Dynamic Spatial Operations	31
2.1.6 Mapping Spatial Objects into Java	32
2.2 Three-Step Database Design	34
2.2.1 The ER Model	35
2.2.2 The Relational Model	37
2.2.3 Mapping the ER Model to the Relational Model	38
2.3 Trends: Extending the ER Model with Spatial Concepts	41
2.3.1 Extending the ER Model with Pictograms	41
2.4 Trends: Object-Oriented Data Modeling with UML	45
2.4.1 Comparison between ER and UML	47
2.5 Summary	48
Bibliographic Notes	49

3	Spatial Query Languages	52
3.1	Standard Database Query Languages	53
3.1.1	World Database	53
3.2	RA	55
3.2.1	The Select and Project Operations	55
3.2.2	Set Operations	56
3.2.3	Join Operation	57
3.3	Basic SQL Primer	59
3.3.1	DDL	59
3.3.2	DML	60
3.3.3	Basic Form of an SQL Query	60
3.3.4	Example Queries in SQL	61
3.3.5	Summary of RA and SQL	64
3.4	Extending SQL for Spatial Data	64
3.4.1	The OGIS Standard for Extending SQL	65
3.4.2	Limitations of the Standard	66
3.5	Example Queries that Emphasize Spatial Aspects	67
3.6	Trends: Object-Relational SQL	71
3.6.1	A Glance at SQL3	72
3.6.2	Object-Relational Schema	73
3.6.3	Example Queries	75
3.7	Summary	75
	Bibliographic Notes	76
3.8	Appendix: <i>State Park Database</i>	79
3.8.1	Example Queries in RA	80
4	Spatial Storage and Indexing	83
4.1	Storage: Disks and Files	84
4.1.1	Disk Geometry and Implications	85
4.1.2	Buffer Manager	86
4.1.3	Field, Record, and File	87
4.1.4	File Structures	88
4.1.5	Clustering	90
4.2	Spatial Indexing	96
4.2.1	Grid Files	97
4.2.2	R-Trees	99
4.2.3	Cost Models	103
4.3	Trends	104
4.3.1	TR*-Tree for Object Decomposition	104
4.3.2	Concurrency Control	105
4.3.3	Spatial Join Index	107
4.4	Summary	111
	Bibliographic Notes	111
5	Query Processing and Optimization	114
5.1	Evaluation of Spatial Operations	115
5.1.1	Overview	115

5.1.2	Spatial Operations	115
5.1.3	Two-Step Query Processing of Object Operations	116
5.1.4	Techniques for Spatial Selection	117
5.1.5	General Spatial Selection	118
5.1.6	Algorithms for Spatial-Join Operations	119
5.1.7	Strategies for Spatial Aggregate Operation: Nearest Neighbor	122
5.2	Query Optimization	122
5.2.1	Logical Transformation	124
5.2.2	Cost-Based Optimization: Dynamic Programming	127
5.3	Analysis of Spatial Index Structures	129
5.3.1	Enumeration of Alternate Plans	131
5.3.2	Decomposition and Merge in Hybrid Architecture	132
5.4	Distributed Spatial Database Systems	132
5.4.1	Distributed DBMS Architecture	134
5.4.2	The Semijoin Operation	135
5.4.3	Web-Based Spatial Database Systems	136
5.5	Parallel Spatial Database Systems	138
5.5.1	Hardware Architectures	139
5.5.2	Parallel Query Evaluation	140
5.5.3	Application: Real-Time Terrain Visualization	142
5.6	Summary	145
	Bibliographic Notes	146
6	Spatial Networks	149
6.1	Example Network Databases	149
6.2	Conceptual, Logical, and Physical Data Models	151
6.2.1	A Logical Data Model	151
6.2.2	Physical Data Models	154
6.3	Query Language for Graphs	157
6.3.1	Shortcomings of RA	158
6.3.2	SQL CONNECT Clause	159
6.3.3	Example Queries on the BART System	161
6.3.4	Trends: SQL3 Recursion	163
6.3.5	Trends: SQL3 ADTs for Networks	164
6.4	Graph Algorithms	165
6.4.1	Path-Query Processing	166
6.4.2	Graph Traversal Algorithms	166
6.4.3	Best-First Algorithm for Single Pair (v, d) Shortest Path	169
6.4.4	Trends: Hierarchical Strategies	170
6.5	Trends: Access Methods for Spatial Networks	173
6.5.1	A Measure of I/O Cost for Network Operations	174
6.5.2	A Graph-Partitioning Approach to Reduce Disk I/O	176
6.5.3	CCAM: A Connectivity Clustered Access Method for Spatial Network	177
6.5.4	Summary	179
	Bibliographic Notes	179

7	Introduction to Spatial Data Mining	182
7.1	Pattern Discovery	183
7.1.1	The Data-Mining Process	183
7.1.2	Statistics and Data Mining	185
7.1.3	Data Mining as a Search Problem	185
7.1.4	Unique Features of Spatial Data Mining	185
7.1.5	Famous Historical Examples of Spatial Data Exploration	186
7.2	Motivating Spatial Data Mining	187
7.2.1	An Illustrative Application Domain	187
7.2.2	Measures of Spatial Form and Auto-correlation	189
7.2.3	Spatial Statistical Models	191
7.2.4	The Data-Mining Trinity	193
7.3	Classification Techniques	194
7.3.1	Linear Regression	195
7.3.2	Spatial Regression	195
7.3.3	Model Evaluation	196
7.3.4	Predicting Location Using Map Similarity (PLUMS)	198
7.3.5	Markov Random Fields	198
7.4	Association Rule Discovery Techniques	202
7.4.1	Apriori: An Algorithm for Calculating Frequent Itemsets	202
7.4.2	Spatial Association Rules	204
7.4.3	Colocation Rules	204
7.5	Clustering	206
7.5.1	K-medoid: An Algorithm for Clustering	209
7.5.2	Clustering, Mixture Analysis, and the EM Algorithm	210
7.5.3	Strategies for Clustering Large Spatial Databases	213
7.6	Spatial Outlier Detection	215
7.7	Summary	221
7.8	Appendix: Bayesian Calculus	221
7.8.1	Conditional Probability	221
7.8.2	Maximum Likelihood	222
	Bibliographic Notes	222
8	Trends in Spatial Databases	227
8.1	Database Support for Field Entities	227
8.1.1	Raster and Image Operations	228
8.1.2	Storage and Indexing	231
8.2	Content-Based Retrieval	233
8.2.1	Topological Similarity	233
8.2.2	Directional Similarity	234
8.2.3	Distance Similarity	235
8.2.4	Attribute Relational Graphs	235
8.2.5	Retrieval Step	237
8.3	Introduction to Spatial Data Warehouses	237
8.3.1	Aggregate Operations	238
8.3.2	An Example of Geometric Aggregation	240

List of Figures

8.3.3 Aggregation Hierarchy 242
 8.3.4 What Is an Aggregation Hierarchy Used For? 245
 8.4 Summary 246
 Bibliographic Notes 246

Bibliography **250**
Index **258**

1.2 Landsat image of Ramsey county, Minnesota, with spatial layers of information superimposed. 6
 1.3 Census blocks with boundary ID:1050. 7
 1.4 Four tables required in a relational database with overlapping attributes to accommodate the polyline datatype. 8
 1.5 Evolution of databases. [Khoshdelian and Baker, 1998] 9
 1.6 Three-layer architecture. 10
 1.7 Two relations to illustrate the difference between join and spatial join. 14
 1.8 The filter-refine strategy for reducing computation time. 15
 1.9 Determining pairs of intersecting rectangles: (a) two sets of rectangles: *R* and *S*; (b) a rectangle *T* with its lower-left and upper-right corners marked; and (c) the sorted set of rectangles with their joins. Note the filtering nature of the plane sweep algorithm. In the example, twelve possible rectangle pairs could be joined. The filtering step reduced the number of possibilities to five. An exact geometry test can be used to check which of the five pairs of objects satisfy the query predicate [Beckmann et al., 1990]. 16
 1.10 (a) Programmer's viewpoint; (b) DBMS designer's viewpoint. 17
 1.11 Different ways of ordering multidimensional data: (a) row order; (b) Z-order. If a line is drawn following the numbers in ascending order, the Z pattern will become obvious. 18
 1.12 (a) Binary tree; (b) B-tree. 18
 1.13 R-tree extends B-tree for spatial objects. 19
 2.1 The object-field dichotomy. (a) A map showing three forest stands: *Pine*, *Fir*, *Oak*. (b) An *object* viewpoint representing the map as a collection of three objects. Each object has a unique identifier, a dominant tree species, and an area. The boundary of the area, a polygon is specified by the coordinates. (c) A *field* viewpoint, where each point is mapped to the value of the dominant tree species. 24
 2.2 An OGIS proposal for building blocks of spatial geometry in UML notation [OGIS, 1999]. 27
 2.3 The nine-intersection model [Egenhofer et al., 1989]. 30
 2.4 An ER diagram for the *State-Park* example. 37
 2.5 Relational schema for the *State-Park* example. 39
 2.6 Schema for point, line, polygon, and elevation. 40
 2.7 ER diagram for the *State-Park* example, with pictograms. 45
 2.8 The *State-Park* example in UML. 46