Contents

Preface VII

List of Figures XV

List of Tables XIX

Part I. FUNDAMENTAL CONCEPTS

1. Introduction 3
1.1 The Original R-tree 7
1.2 Summary 13

2. Dynamic Versions of R-trees 15
2.1 The R^+-tree 15
2.2 The R*-tree 18
2.3 The Hilbert R-tree 20
2.4 Linear Node Splitting 22
2.5 Optimal Node Splitting 24
2.6 Branch Grafting 25
2.7 Compact R-trees 27
2.8 cR-trees 27
2.9 Deviating Variations 29
2.9.1 PR-trees 30
2.9.2 LR-trees 31
2.10 Summary 34

3. Static Versions of R-trees 35
3.1 The Packed R-tree 35
3.2 The Hilbert Packed R-tree 36
3.3 The STR R-tree 37
3.4 Top-Down Packing Techniques 38
3.5 Small-Tree-Large-Tree and GBI 40
3.6 Bulk Insertion by Seeded Clustering 42
3.7 The Buffer R-tree 44
3.8 R-tree with Low Stabbing Number 45

3.9 Merging R-trees 45
3.10 Summary 47

Part II. QUERY PROCESSING ISSUES

4. Fundamental Query Processing Techniques 51
4.1 Two-step Processing 51
4.2 Range and Topological Queries 53
4.3 Nearest-Neighbor Queries 55
4.3.1 A Branch-and-Bound Algorithm 56
4.3.2 An Improvement to the Original Algorithm 58
4.3.3 Incremental Nearest-Neighbor Searching 59
4.3.4 Comparison of Nearest Neighbor Algorithms 61
4.4 Spatial Join Queries 62
4.4.1 Algorithm Based on Depth-First Traversal 62
4.4.2 Algorithm Based on Breadth-First Traversal 65
4.4.3 Join Between an R-tree-Indexed and a Non-Indexed Dataset 67
4.5 Summary 68

5. Processing More Complex Queries 69
5.1 Categorical Range Queries 69
5.2 Reverse and Constrained Nearest-Neighbor Queries 72
5.2.1 Reverse Nearest Neighbors 72
5.2.2 Generalized Constrained Nearest Neighbor Searching . . 75
5.3 Multi-way Spatial Join Queries 77
5.4 Incremental Distance-Join and Closest-Pair Queries 80
5.4.1 Incremental Distance Join 80
5.4.2 Distance Semi-Join Query 83
5.4.3 Finding Closest Pairs 83
5.5 All Nearest-Neighbor Queries 85
5.6 Approximate Query Processing on R-trees 87
5.7 Classification of R-tree-Based Query Processing Algorithms . . . 93
5.8 Summary 94

Part III. R-TREES IN MODERN APPLICATIONS

6. R-trees in Spatiotemporal Databases 99
6.1 Preliminaries 99
6.2 The RT-tree 101
6.3 The 3D R-tree 101
6.4 The 2+3 R-tree 102
6.5 The Historical R-tree 103
6.6 The R^{ST}-tree 104

6.7 The Partially Persistent R-tree . . . 105
6.8 The MV3R-tree . . . 106
6.9 The TB-tree . . . 108
6.10 Scalable and Efficient Trajectory Index (SETI) . . . 109
6.11 The Q+R-tree . . . 110
6.12 The FNR-tree and the MON-tree . . . 111
6.13 The Time-Parameterized R-tree . . . 112
6.14 The VCI R-tree . . . 114
6.15 Summary . . . 115

7. R-trees for Multimedia, Warehousing and Mining . . . 117
7.1 R-trees in Multimedia Databases . . . 117
7.1.1 Generic Multimedia Indexing (GEMINI) . . . 117
7.1.2 High-Dimensional Access Methods . . . 121
7.1.3 R-trees and Hidden Markov Models in Music Retrieval . . . 125
7.1.4 R-trees and Self-Organizing Maps . . . 126
7.2 R-trees in Data Warehousing and Data Mining . . . 126
7.3 Summary . . . 130

Part IV. ADVANCED ISSUES

8. Query Optimization Issues . . . 133
8.1 Selectivity and Cost Models for Selection Queries . . . 133
8.1.1 Formulae for Range Queries . . . 133
8.1.2 Formulae for Nearest-Neighbor Queries . . . 140
8.2 Selectivity and Cost Models for Join Queries . . . 142
8.2.1 Formulae for Pair-Wise Joins . . . 142
8.2.2 Formulae for Multiway Joins . . . 144
8.2.3 Formulae for Distance-Join Queries . . . 146
8.3 Spatiotemporal Query Optimization . . . 147
8.4 Sampling and Histogram-Based Techniques . . . 149
8.5 Summary . . . 150

9. Implementation Issues . . . 151
9.1 Parallel Systems . . . 151
9.1.1 Multidisk Systems . . . 152
9.1.2 Multiprocessor Systems . . . 156
9.2 Concurrency Control . . . 159
9.2.1 R-link Method . . . 160
9.2.2 Top-down Approaches . . . 161
9.3 Issues in Relational Implementations . . . 162
9.3.1 Stochastic Driven Relational R-trees . . . 162
9.3.2 Lazy Deletion Methods . . . 164
9.3.3 R-trees in Research Prototypes . . . 165

9.3.4 R-trees in Commercial Database Systems 169
9.4 Summary 171

Epilogue 173

References 175

Index 191