

Contents

1	Advanced Query Processing: An Introduction	1
	<i>Barbara Catania, Lakhmi Jain</i>	
1.1	Introduction	1
1.2	Preference-Based Query Processing	5
1.3	Approximate Query Processing for Non-traditional Data	6
1.4	Continuous Query Processing	8
1.5	Adaptive Query Processing	9
1.6	Queries over Heterogeneous Data Sources	10
1.7	Conclusion and Discussion	11
	References	12

Part I

2	On Skyline Queries and How to Choose from Pareto Sets	15
	<i>Christoph Lofi, Wolf-Tilo Balke</i>	
2.1	Introduction	15
2.2	Formalization of Skyline Sets Following the Pareto Semantics	17
2.3	Relaxing the Pareto Semantics	21
2.4	Summarizing the Skyline	23
2.4.1	Approximately Dominating Representatives	23
2.4.2	Statistical Sampling Skylines	24
2.5	Weighting Characteristics of Skyline Points	25
2.5.1	Skycubes and Subspace Analysis	26
2.5.2	SKYRANK	28
2.5.3	k Most Representative Skyline Points	29
2.5.4	Personalized Top-k Retrieval / Telescope	30
2.6	Cooperative Approaches	31
2.6.1	Interactive Preference Elicitation	31
2.6.2	Trade-Off Skylines	32
2.7	Conclusion and Discussion	33
	References	34

3 Processing Framework for Ranking and Skyline Queries 37
Seung-won Hwang

3.1 Introduction 37

3.2 Related Work 40

3.2.1 Sorted-Order Algorithms 40

3.2.2 Partitioning-Based Algorithms 41

3.2.3 Taxonomy and Generalization 41

3.3 Framework for Ranking Queries 42

3.3.1 Preliminaries 42

3.3.2 Extracting Skeleton 44

3.3.3 Cost-Based Optimization Strategies 45

3.4 Framework for Skyline Queries 47

3.4.1 Preliminaries 47

3.4.2 Extracting Skeleton 49

3.4.3 Cost-Based Optimization Strategies 53

3.5 Conclusion and Open Issues 55

References 55

4 Preference-Based Query Personalization 57
Georgia Koutrika, Evaggelia Pitoura, Kostas Stefanidis

4.1 Introduction 57

4.2 Preference Representation 59

4.2.1 Context Specification 59

4.2.2 Preference Specification 61

4.2.3 Combining Preferences 64

4.2.4 Example: A User Profile 66

4.3 Personalizing Queries Using Preferences 68

4.3.1 Preference Selection 68

4.3.2 Personalized Query Processing 73

4.4 Preference Learning 76

4.5 Conclusion and Open Issues 77

References 78

Part II

5 Approximate Queries for Spatial Data 83
Alberto Belussi, Barbara Catania, Sara Migliorini

5.1 Introduction 83

5.2 Background on Spatial Data and Queries 86

5.3 A Taxonomy of Query-Based Approximation Techniques for
Spatial Data 89

5.3.1 Query Relaxation 89

5.3.2 Approximate Query Processing 92

5.4	Spatial Top- k Queries	93
5.4.1	Top- k Ranking Function	94
5.4.2	Top- k Query Processing Algorithms	96
5.5	Spatial Skyline Queries	104
5.5.1	Spatial Skyline Queries	105
5.5.2	Spatial Skyline Query Processing Algorithms	106
5.6	Approximate Query Processing	112
5.6.1	Approximate Algorithms for Multiway Spatial Join	113
5.6.2	Approximate Algorithms for Distance-Based Queries	115
5.6.3	Algorithms Based on Approximate Spatial Data	116
5.7	Towards Qualitative Approximation Techniques for Spatial Data	118
5.7.1	From Qualitative to Quantitative Spatial Relations	119
5.7.2	Spatial Top- k Queries Based on Qualitative Relations	121
5.7.3	Spatial Skyline Queries Based on Qualitative Relations	123
5.8	Conclusion and Open Issues	124
	References	124
6	Approximate XML Query Processing	129
	<i>Giovanna Guerrini</i>	
6.1	Introduction	129
6.2	Twig Queries	132
6.2.1	XML Documents	133
6.2.2	Twig Queries: Definition	134
6.2.3	(Exact) Twig Query Processing	135
6.2.4	Twig Queries as a Basis for Approximate Querying	136
6.3	Various Extents of Approximation	138
6.3.1	Vocabulary	138
6.3.2	Hierarchical Structure	140
6.4	Ranking	142
6.4.1	Tree Edit Distance	143
6.4.2	An Alternative Match Based Similarity	144
6.4.3	Structure and Content $tf \cdot idf$ Scoring	145
6.4.4	Content Scoring with Structure Filters	146
6.5	Approximate Query Processing	148
6.5.1	Twig-Path Scoring and Whirpool	148
6.5.2	TopX	149

6.5.3	TASM	150
6.5.4	ArHeX	150
6.6	Conclusion and Discussion	151
	References	153

Part III

7	Progressive and Approximate Join Algorithms on Data Streams ...	157
	<i>Wee Hyong Tok, Stéphane Bressan</i>	
7.1	Introduction	157
7.2	Background	158
7.3	Why Progressive Joins?	160
7.4	Joins from Different Data Models Flock Together	161
7.4.1	Relational Joins	161
7.4.2	Spatial Joins	163
7.4.3	High-Dimensional Distance-Similarity Joins	164
7.4.4	Progressive XML Structural Joins	164
7.5	Generic Progressive Join Framework	164
7.5.1	Building Blocks for Generic Progressive Join Framework	165
7.5.2	Progressive Join Framework	166
7.5.3	RRPJ Instantiations	170
7.6	Progressive Approximate Joins	170
7.6.1	Extreme Scenario	171
7.6.2	Measuring the Performance of Progressive, Approximate Joins	172
7.6.3	Different Types of Progressive, Approximate Joins	174
7.6.4	Discussion	180
7.7	Open Issues	181
7.8	Conclusion	182
	References	183
8	Online Aggregation	187
	<i>Sai Wu, Beng Chin Ooi, Kian-Lee Tan</i>	
8.1	Introduction	187
8.2	Basic Principles	190
8.2.1	Statistical Model	192
8.2.2	Sampling	193
8.3	Advanced Applications	194
8.3.1	Online Aggregation for Multi-relation Query Processing	194
8.3.2	Online Aggregation for Multi-query Processing	197

8.3.3	Distributed Online Aggregation	203
8.3.4	Online Aggregation and MapReduce	207
8.4	Conclusion and Discussion	209
	References	209

Part IV

9	Adaptive Query Processing in Distributed Settings	211
	<i>Anastasios Gounaris, Efthymia Tsamoura, Yannis Manolopoulos</i>	
9.1	Introduction	211
9.1.1	Distributed Query Processing Basics	213
9.1.2	Related Work	214
9.2	A Framework for Analysis of AdQP	214
9.3	AdQP in Centralized Settings	215
9.3.1	Overview of Techniques	215
9.3.2	On Applying Conventional AdQP Techniques in Distributed Settings	216
9.4	AdQP for Distributed Settings: Extensions to Eddies	217
9.4.1	Techniques	218
9.4.2	Summary	220
9.5	AdQP for Distributed Settings: Operator Load Management	220
9.5.1	Intra-Operator Load Management	221
9.5.2	Inter-Operator Load Management	226
9.5.3	More Generic Solutions	230
9.6	AdQP for Distributed Settings: Other Techniques	230
9.7	Conclusion and Open Issues	232
	References	233
10	Approximate Queries with Adaptive Processing	237
	<i>Barbara Catania, Giovanna Guerrini</i>	
10.1	Introduction	237
10.2	QoD ² Techniques: Some Examples	241
10.2.1	Adaptively Approximate Pipelined Joins	241
10.2.2	Adaptive Processing of Skyline-Based Queries over Data Streams	243
10.3	QoD-Oriented Approximate Queries	244
10.3.1	Query Rewriting	245
10.3.2	Preference-Based Queries	246
10.3.3	Approximate Query Processing	248
10.4	QoS-Oriented Approximate Queries	249
10.4.1	Data Reduction	250
10.4.2	Load Shedding	253
10.4.3	Approximation of the Processing Algorithm	254

10.5 Adaptive Query Processing 254

10.5.1 Adapting Query Plans 255

10.5.2 Adaptively Coping with Limited Resources under
Fixed Plans 259

10.5.3 Further Adaptation Subjects 260

10.6 Conclusion and Discussion 262

References 263

Part V

11 Querying Conflicting Web Data Sources 271

*Gilles Nachouki, Mohamed Quafafou, Omar Boucelma,
François-Marie Colonna*

11.1 Introduction 272

11.2 Conflicting Web Data Sources 274

11.2.1 Overview of Conflict Types 274

11.2.2 Conflicting Data in Life Sciences 275

11.2.3 Assumptions about Conflict Representation 277

11.3 Mediating Biological Conflicting Data with BGLAV 277

11.3.1 BGLAV Overview 277

11.3.2 Query Processing in BGLAV 278

11.4 MFA - Multi-source Fusion Approach 281

11.4.1 MFA Overview 281

11.4.2 Methodology for Semantic Reconciliation 285

11.4.3 Query Processing in MFA 287

11.5 Application 294

11.5.1 Data Source Description 295

11.5.2 BGLAV Illustrating Examples 295

11.5.3 MFA Illustrating Examples 297

11.5.4 Evaluation of MFA Queries 300

11.6 Conclusion and Open Issues 301

References 301

12 A Functional Model for Dataspace Management Systems 305

*Cornelia Hedeler, Alvaro A.A. Fernandes, Khalid Belhajjame,
Lu Mao, Chenjuan Guo, Norman W. Paton, Suzanne M. Embury*

12.1 Introduction 305

12.2 Dataspace Life Cycle 307

12.3 Background 309

12.3.1 Dataspace Management Systems 310

12.3.2 Model Management Systems 312

12.4 Functional Model 313

12.4.1 An Overview 313

12.4.2 Preliminary Assumptions 314

12.4.3 Intensional Descriptions 315

12.4.4 Sorts 318

- 12.4.5 Operations 321
- 12.5 Bioinformatics Use Case 329
 - 12.5.1 Example: Dataspace Initialization 329
 - 12.5.2 Example: Dataspace Maintenance 331
 - 12.5.3 Example: Dataspace Improvement 334
- 12.6 Conclusion and Open Issues 336
- References 337
- Author Index 343**
- Subject Index 345**
- Editors 349**