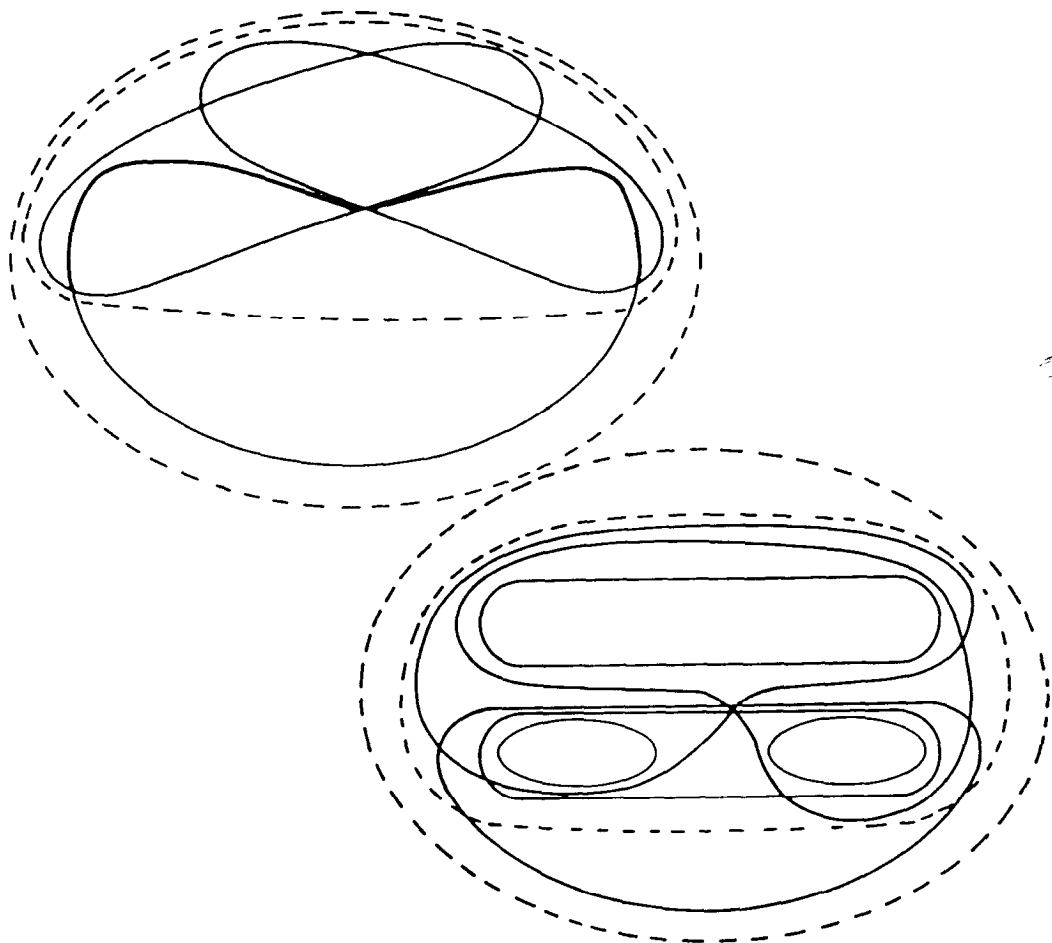


THE THEORY OF RELATIONAL DATABASES

DAVID MAIER



COMPUTER SCIENCE PRESS

TABLE OF CONTENTS

Preface	xiv
Acknowledgements	xv
1. RELATIONS AND RELATION SCHEMES	1
1.1. Brass Tacks	1
1.2. Formalization of Relations	2
1.3. Keys	4
1.4. Updates to Relations	5
1.5. Exercises	8
1.6. Bibliography and Comments	10
2. RELATIONAL OPERATORS	11
2.1. Boolean Operations	11
2.2. The Select Operator	13
2.3. The Project Operator	15
2.4. The Join Operator	16
2.5. Properties of Join	18
2.6. Exercises	22
2.7. Bibliography and Comments	24
3. MORE OPERATIONS ON RELATIONS	25
3.1. The Divide Operator	25
3.2. Constant Relations	26
3.3. Renaming Attributes	27
3.4. The Equijoin Operator	29
3.5. Extensions for Other Comparisons on Domains	31
3.5.1. Extending Selection	32
3.5.2. The Theta-Join Operator	33
3.6. Relational Algebra	34
3.6.1. Algebraic Expressions as Mappings	35
3.6.2. Restricting the Set of Operators	36

3.7.	The Split Operator	37
3.8.	The Factor Operator	38
3.9.	Exercises	39
3.10.	Bibliography and Comments	41
4.	FUNCTIONAL DEPENDENCIES	42
4.1.	Definitions	42
4.2.	Inference Axioms	44
4.3.	Applying the Inference Axioms	47
4.4.	Completeness of the Inference Axioms	49
4.5.	Derivations and Derivation DAGs	51
4.5.1.	RAP-Derivation Sequences	53
4.5.2.	Derivation DAGs	56
4.5.3.	More about Derivation DAGs	60
4.6.	Testing Membership in F^+	63
4.7.	Exercises	69
4.8.	Bibliography and Comments	70
5.	COVERS FOR FUNCTIONAL DEPENDENCIES	71
5.1.	Covers and Equivalence	71
5.2.	Nonredundant Covers	72
5.3.	Extraneous Attributes	74
5.4.	Canonical Covers	77
5.5.	The Structure of Nonredundant Covers	78
5.6.	Minimum Covers	79
5.6.1.	Direct Determination	79
5.6.2.	Computing Minimum Covers	84
5.7.	Optimal Covers	86
5.8.	Annular Covers and Compound Functional Dependencies	87
5.9.	Exercises	90
5.10.	Bibliography and Comments	92
6.	DATABASES AND NORMAL FORMS	93
6.1.	Databases and Database Schemes	94
6.2.	Normal Forms for Databases	96
6.2.1.	First Normal Form	96
6.2.2.	Anomalies and Data Redundancy	98
6.2.3.	Second Normal Form	99
6.2.4.	Third Normal Form	99
6.3.	Normalization through Decomposition	101

6.4.	Shortcomings of Normalization through Decomposition . . .	104
6.5.	Normalization through Synthesis	107
6.5.1.	Preliminary Results for the Synthesis Algorithm . . .	108
6.5.2.	Developing the Synthesis Algorithm	108
6.5.3.	Correctness and Other Properties of the Synthesis Algorithm	110
6.5.4.	Refinements of the Synthesis Algorithm	113
6.6.	Avoidable Attributes	115
6.7.	Boyce-Codd Normal Form	117
6.7.1.	Problems with Boyce-Codd Normal Form	119
6.8.	Exercises	119
6.9.	Bibliography and Comments	122
7.	MULTIVALUED DEPENDENCIES, JOIN DEPENDENCIES, AND FURTHER NORMAL FORMS	123
7.1.	Multivalued Dependencies	124
7.2.	Properties of Multivalued Dependencies	126
7.3.	Multivalued Dependencies and Functional Dependencies	127
7.4.	Inference Axioms for Multivalued Dependencies	129
7.4.1.	Multivalued Dependencies Alone	129
7.4.2.	Functional and Multivalued Dependencies	132
7.4.3.	Completeness of the Axioms and Computing Implications	133
7.5.	Fourth Normal Form	135
7.6.	Fourth Normal Form and Enforceability of Dependencies	137
7.7.	Join Dependencies	139
7.8.	Project-Join Normal Form	140
7.9.	Embedded Join Dependencies	142
7.10.	Exercises	143
7.11.	Bibliography and Comments	144
8.	PROJECT-JOIN MAPPINGS, TABLEAUX, AND THE CHASE	146
8.1.	Project-Join Mappings	146
8.2.	Tableaux	148
8.2.1.	Tableaux as Mappings	150
8.2.2.	Representing Project-Join Mappings as Tableaux	151
8.3.	Tableaux Equivalence and Scheme Equivalence	152

8.4.	Containment Mappings	156
8.5.	Equivalence with Constraints	160
	8.5.1. F-rules	162
	8.5.2. J-rules	163
8.6.	The Chase	164
	8.6.1. The Finite Church-Rosser Property	168
	8.6.2. Equivalence of Tableaux under Constraints	174
	8.6.3. Testing Implication of Join Dependencies	175
	8.6.4. Testing Implication of Functional Dependencies ..	177
	8.6.5. Computing a Dependency Basis	180
8.7.	Tableaux as Templates	182
8.8.	Computational Properties of the Chase Computation	186
8.9.	Exercises	189
8.10.	Bibliography and Comments	194
9.	REPRESENTATION THEORY	195
9.1.	Notions of Adequate Representation	195
9.2.	Data-Equivalence of Database Schemes	208
9.3.	Testing Adequate Representation and Equivalence Under Constraints	210
	9.3.1. P Specified by Functional Dependencies Only	211
	9.3.2. P Specified by Functional and Multivalued Dependencies	215
	9.3.3. Testing Data-Equivalence	217
9.4.	Exercises	221
9.5.	Bibliography and Comments	223
10.	QUERY SYSTEMS	224
10.1.	Equivalence and Completeness	225
10.2.	Tuple Relational Calculus	227
	10.2.1. Tuple Calculus Formulas	229
	10.2.2. Types, and Free and Bound Occurrences	231
	10.2.3. Tuple Calculus Expressions	236
10.3.	Reducing Relational Algebra with Complement to Tuple Relational Calculus	242
10.4.	Limited Interpretation of Tuple Calculus Formulas	244
	10.4.1. Reducing Relational Algebra to Tuple Calculus with Limited Evaluation	247
	10.4.2. Safe Tuple Calculus Expressions	247
10.5.	Domain Relational Calculus	250
10.6.	Reduction of Tuple Calculus to Domain Calculus	255

10.7.	Reduction of Domain Calculus to Relational Algebra	257
10.8.	Tableau Queries	262
10.8.1.	Single Relation Tableau Queries	262
10.8.2.	Tableau Queries for Restricted Algebraic Expressions	268
10.8.3.	Tableau Queries that Come from Algebraic Expressions	272
10.8.4.	Tableau Queries for Multirelation Databases	274
10.8.5.	Tableau Set Queries	276
10.9.	Conjunctive Queries	278
10.10.	Exercises	278
10.11.	Bibliography and Comments	286
11.	QUERY MODIFICATION	287
11.1.	Levels of Information in Query Modification	293
11.2.	Simplifications and Common Subexpressions in Algebraic Expressions	295
11.3.	Optimizing Algebraic Expressions	301
11.4.	Query Decomposition	307
11.4.1.	Instantiation	311
11.4.2.	Iteration	313
11.4.3.	The Query Decomposition Algorithm	315
11.5.	Tableau Query Optimization	323
11.5.1.	Tableau Query Equivalence	323
11.5.2.	Simple Tableau Queries	327
11.5.3.	Equivalence with Constraints	335
11.5.4.	Extensions for Multiple-Relation Databases	339
11.5.5.	Tableau Set Query Equivalence	348
11.6.	Optimizing Conjunctive Queries	350
11.7.	Query Modification for Distributed Databases	353
11.7.1.	Semijoins	354
11.7.2.	Fragments of Relations	359
11.8.	Exercises	361
11.9.	Bibliography and Index	369
12.	NULL VALUES, PARTIAL INFORMATION AND DATABASES SEMANTICS	371
12.1.	Nulls	372
12.2.	Functional Dependencies and Nulls	377
12.3.	Constraints on Nulls	384
12.4.	Relational Algebra and Partial Relations	386

12.4.1.	Possibility Functions	386
12.4.2.	Generalizing the Relational Operators	389
12.4.3.	Specific Possibility Functions	394
12.5.	Partial Information and Database Semantics	406
12.5.1.	Universal Relation Assumptions	406
12.5.2.	Placeholders and Subscheme Relations	408
12.5.3.	Database Semantics and Window Functions	410
12.5.4.	A Window Function Based on Joins	413
12.5.5.	Weak Instances	416
12.5.6.	Independence	422
12.5.7.	A Further Condition on Window Functions	427
12.6.	Exercises	432
12.7.	Bibliography and Comments	437
13.	ACYCYLIC DATABASE SCHEMES	439
13.1.	Properties of Database Schemes	439
13.1.1.	Existence of a Full Reducer	439
13.1.2.	Equivalence of a Join Dependency to Multivalued Dependencies	442
13.1.3.	Unique 4NF Decomposition	443
13.1.4.	Pairwise Consistency Implies Total Consistency	444
13.1.5.	Small Intermediate Joins	445
13.2.	Syntactic Conditions on Database Schemes	447
13.2.1.	Acyclic Hypergraphs	447
13.2.2.	Join Trees	452
13.2.3.	The Running Intersection Property	455
13.3.	Equivalence of Conditions	455
13.3.1.	Graham Reduction	456
13.3.2.	Finding Join Trees	457
13.3.3.	The Equivalence Theorem for Acyclic Database Schemes	460
13.3.4.	Conclusions	477
13.4.	Exercises	478
13.5.	Bibliography and Comments	482
14.	ASSORTED TOPICS	485
14.1.	Logic and Data Dependencies	485
14.1.1.	The World of Two-Tuple Relations	486
14.1.2.	Equivalence of Implication for Logic and Functional Dependencies	488

14.1.3.	Adding Multivalued Dependencies	489
14.1.4.	Nonextendability of Results	492
14.2.	More Data Dependencies	493
14.2.1.	Template Dependencies	494
14.2.2.	Examples and Counterexamples for Template Dependencies	498
14.2.3.	A Graphical Representation for Template Dependencies	500
14.2.4.	Testing Implication of Template Dependencies	506
14.2.5.	Generalized Functional Dependencies	516
14.2.6.	Closure of Satisfaction Classes Under Projection	524
14.3.	Limitations of Relational Algebra	527
14.4.	Computed Relations	533
14.4.1.	An Example	533
14.4.2.	Testing Expressions Containing Computed Relations	536
14.5.	Exercises	542
14.6.	Bibliography and Comments	547
15.	RELATIONAL QUERY LANGUAGES	550
15.1.	ISBL	551
15.2.	QUEL	556
15.3.	SQL	561
15.4.	QBE	568
15.5.	PIQUE	583
15.6.	Bibliography and Comments	591
BIBLIOGRAPHY		593
INDEX		611

ABOUT THE AUTHOR

David Maier received his BA degree in mathematics and computer science from the University of Oregon in 1974 and his PhD from Princeton University in 1978. For four years he was assistant professor of computer science with the State University of New York at Stony Brook. He is currently assistant professor of computer science at the Oregon Graduate Center.

Dr. Maier's papers on database theory have appeared in *JACM*, *ACM Transactions on Database Systems*, and the *SIAM Journal of Computing*.

ABOUT THE BOOK

This remarkably comprehensive new book assembles concepts and results in relational databases theory previously scattered through journals, books, conference proceedings, and technical memoranda in one convenient source, and introduces pertinent new material not found elsewhere. The book is intended for a second course in databases, but is an excellent reference for researchers in the field. The material covered includes relational algebra, functional dependencies, multivalued and join dependencies, normal forms, tableaux and the chase computation, representation theory, domain and tuple relational calculus, query modification, database semantics and null values, acyclic database schemes, template dependencies, and computed relations. The final chapter is a brief survey of query languages in existing relational systems. Each chapter contains numerous examples and exercises, along with bibliographic remarks.