

Constraint Databases

Gabriel Kuper • Leonid Libkin
Jan Paredaens (Eds.)

Constraint Databases

With 83 Figures



Springer

Editors

Gabriel Kuper
Leonid Libkin

Bell Labs/Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974, USA
{kuper, libkin}@bell-labs.com

Jan Paredaens

Department of Mathematics and
Computer Science
University of Antwerp (UIA)
Universiteitsplein 1
2610 Wilrijk-Antwerp, Belgium
pareda@uia.ua.ac.be

Library of Congress Cataloging-in-Publication Data

Kuper, Gabriel.

Constraint databases/Gabriel Kuper, Leonid Libkin, Jan Paredaens (eds).
p. cm.

Includes bibliographical references and index.

ISBN 978-3-642-08542-0

ISBN 978-3-662-04031-7 (eBook)

DOI 10.1007/978-3-662-04031-7

1. Constraint databases. I. Libkin, Leonid. II. Paredaens, Jan, 1947- III. Title.

QA76.9.C67 K87 2000

005.74-dc21

00-021716

ACM Computing Classification (1998): H.1-2, D.3.2, F.4.1

ISBN 978-3-642-08542-0

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

© Springer-Verlag Berlin Heidelberg 2000

Originally published by Springer-Verlag Berlin Heidelberg New York in 2000

Softcover reprint of the hardcover 1st edition 2000

The use of general descriptive names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover Design: Künkel + Lopka, Werbeagentur, Heidelberg

Typesetting by the editors using a Springer T_EX macro-package

SPIN 10694592 45/3142/SR - 5 4 3 2 1 0 - Printed on acid-free paper

Preface

This book is the first comprehensive survey of the field of constraint databases. Constraint databases are a fairly new and active area of database research. The key idea is that constraints, such as linear or polynomial equations, are used to represent large, or even infinite, sets in a compact way. The ability to deal with infinite sets makes constraint databases particularly promising as a technology for integrating spatial and temporal data with standard relational databases. Constraint databases bring techniques from a variety of fields, such as logic and model theory, algebraic and computational geometry, as well as symbolic computation, to the design and analysis of data models and query languages.

The book is a collaborative effort involving many authors who have contributed chapters on their fields of expertise. Despite this, the book is designed to be read as a whole, as opposed to a collection of individual surveys. In particular, the terminology and the style of presentation have been standardized, and there are multiple cross-references between the chapters.

The idea of constraint databases goes back to the late Paris Kanellakis. While visiting the IBM T. J. Watson Research Center in 1988, he was shown a demonstration of the CLP(R) system. This system is an extension of logic programming in which rules may contain constraints over some domain, such as real or rational numbers with various arithmetic operations. Paris immediately wondered whether a database theory could be developed for such systems, analogous to the way deductive databases were inspired by logic programming. The direct consequence of this idea was his collaborative work on Constraint Query Languages [KKR90], which started the field of constraint databases. Over the years, Paris proposed many research directions that helped shape the field.

Many people helped us while we were working on the book. Erich Grädel used an early draft for a course he taught at RWTH Aachen, and many people from his Group on Mathematical Foundations of Computer Science provided useful feedback. We particularly want to thank Stephan Kreutzer, Martin Otto, Eric Rosen, and Marc Spielmann. All chapters were refereed; we thank Serge Abiteboul, Foto Afrati, Catriel Beeri, Stavros Cosmadakis, Guozhu Dong, Martin Grohe, Oliver Günther, Manolis Koubarakis, Michael Maher, Frank Neven, Kyuosek Shim, Dan Suciu, Moshe Vardi, Jeffrey Vitter, and

Limsoon Wong for their comments. Several chapters were refereed by book contributors.

We want to thank Bill Coughran, Narain Gehani, and Mihalis Yannakakis of Bell Labs and Serge Abiteboul and Sophie Cluet of INRIA for their support of the project. We thank all members of the Database Systems and Computing Principles Research departments at Bell Labs, the Advanced Database Research and Modeling group at the University of Antwerp, and the Verso group at INRIA for providing a stimulating environment and technical support. We also would like to thank Bart Kuijpers for handling the bibliography, and Hans Wössner from Springer-Verlag for editorial assistance.

Murray Hill, NJ, USA
Murray Hill, NJ, USA and Paris, France
Antwerp, Belgium
January 2000

Gabriel Kuper
Leonid Libkin
Jan Paredaens

Table of Contents

1. Introduction	1
Gabriel Kuper, Leonid Libkin, and Jan Paredaens	
1.1 Motivation and Framework	1
1.2 Relational Databases and First-Order Query Languages	2
1.3 Spatial Data	4
1.4 Constraint Databases	7
1.5 The CDB Model	8
1.6 Querying Constraint Databases	10
1.7 Applications	13
1.8 Historical Note	15

Part I. Theoretical Foundations

2. Constraint Databases, Queries, and Query Languages	21
Jan Van den Bussche	
2.1 Introduction	21
2.2 Logic	21
2.2.1 Quantifier Elimination	23
2.3 The Constraint Database Model	23
2.3.1 Constraints	23
2.3.2 Constraint Databases	25
2.3.3 Testing Equality of Constraint Relations	27
2.4 Queries on Constraint Databases	28
2.4.1 Constraint Queries	28
2.4.2 Relational Calculus with Constraints	30
2.4.3 Computational Feasibility	33
2.4.4 Relational Algebra with Constraints	34
2.5 Computationally Complete Constraint Query Languages	35
2.6 Equivalence and Satisfiability	39
2.7 Conjunctive Queries with Constraints	43
2.8 DATALOG with Constraints	47

2.8.1	Adding Negation	51
2.9	Bibliographic Notes	53
3.	Expressive Power: The Finite Case	55
	Michael Benedikt and Leonid Libkin	
3.1	Introduction	55
3.1.1	Semantics of Constraint Queries	56
3.1.2	Collapse Results	57
3.2	Notation	60
3.2.1	Relational Databases over Infinite Structures	60
3.2.2	First-Order Logic	61
3.2.3	Genericity	62
3.3	Active Semantics	62
3.4	Natural Semantics	66
3.4.1	Natural-Active Collapse	66
3.4.2	O-Minimality	68
3.4.3	Natural-Active Collapse: Algorithm and Proof	69
3.4.4	When the Collapse Fails	73
3.5	Higher-Order Logics	74
3.5.1	Natural Semantics and Hybrid Logics	77
3.6	Other Techniques and Extensions	79
3.7	Conclusion	84
3.8	Bibliographic Notes	84
4.	Expressive Power: The Infinite Case	89
	Stéphane Grumbach, Gabriel Kuper, and Jianwen Su	
4.1	Introduction	89
4.2	Complexity of First-Order Queries	89
4.2.1	FO + POLY	90
4.2.2	Encoding of Boolean Circuits	92
4.2.3	FO + LIN	94
4.2.4	FO + LIN over Restricted Databases	97
4.2.5	FO(<)	98
4.3	Expressive Power of First-Order Queries	98
4.4	Expressive Power of Recursive Languages	102
4.4.1	DATALOG [∇] (<)	102
4.4.2	DATALOG + LIN	104
4.5	Bibliographic Notes	106

5. Query Safety with Constraints	109
Michael Benedikt and Leonid Libkin	
5.1 Introduction	109
5.2 Safe Constraint Queries: The Finite Case	111
5.2.1 Preliminaries	111
5.2.2 Safe Translations	112
5.2.3 Range-Restriction and Safety	114
5.2.4 Deciding Safety	118
5.2.5 Dichotomy Theorem and Outputs of Queries	120
5.3 Safe Constraint Queries: The Infinite Case	121
5.3.1 Preserving Geometric Properties	121
5.3.2 Safety via Coding	122
5.3.3 Examples of Coding	123
5.3.4 Decidability Results and Geometric Bounds	126
5.4 Bibliographic Notes	127
 6. Aggregate Languages for Constraint Databases	 131
Jan Chomicki and Leonid Libkin	
6.1 Introduction	131
6.2 Relational and Spatial Aggregation	132
6.3 Approximating the Volume	136
6.4 Definability of Volume Operators	140
6.5 Restricted Aggregate Operators	144
6.5.1 Variable Independence and Closure	145
6.5.2 FO + POLY + SUM and Volumes of Semi-Linear Sets ..	148
6.6 Conclusion	152
6.7 Bibliographic Notes	153
 7. DATALOG and Constraints	 155
Peter Z. Revesz	
7.1 Introduction	155
7.2 Evaluation of DATALOG with Constraints	155
7.3 Termination, Safety, and Data Complexity	156
7.4 DATALOG with Dense Order Constraints	156
7.5 DATALOG with Gap-Order Constraints	158
7.5.1 General Theory	158
7.5.2 Stratified DATALOG with Gap-Order Constraints	159
7.5.3 DATALOG with Unrestricted Gap-Order Constraints ..	162
7.6 DATALOG with Linear Constraints	163
7.7 DATALOG with Polynomial Constraints	163
7.8 DATALOG with Boolean Equality Constraints	167
7.8.1 General Theory	167

7.8.2 Application: Adder Circuit 168
7.9 Bibliographic Notes 169

Part II. Spatial and Temporal Data

8. Geographic Information Systems 175
Gabriel Kuper and Michel Scholl

8.1 Introduction 175
8.1.1 What is Geographic Information? 175
8.1.2 Are GIS DBMS for Geographic Information?..... 176
8.1.3 Geographic Data Sources 176
8.1.4 GIS Applications 177
8.1.5 Which Spatial Operations are Expected From a GIS? . 177
8.1.6 Brief History of GIS 179
8.2 GIS Data Models 180
8.2.1 Vector vs. Raster Data 181
8.2.2 Spatial Models and Representation of Topology 182
8.2.3 Sample Spatial Database Schema 185
8.2.4 Examples of Queries 186
8.2.5 Limitations of the Current Models 187
8.3 The Constraint Approach for GIS 187
8.4 Representing Spatial Data with Constraints 188
8.4.1 Conversion 188
8.4.2 Storage 190
8.4.3 The Constraint vs. the Vector Approach..... 191
8.5 Queries over Constraint Databases 192
8.6 Role of Constraints in a GIS 196
8.7 Bibliographic Notes 197

9. Linear-Constraint Databases 199
Marc Gyssens, Luc Vandeuren, and Dirk Van Gucht

9.1 Introduction 199
9.2 Properties of Semi-Linear Sets 200
9.3 Positive Expressiveness Results 201
9.4 Negative Expressiveness Results 208
9.5 Extensions of FO + LIN 212
9.5.1 Extensions of FO + LIN with Operators 213
9.5.2 Extension of FO + LIN with Product Variables 214
9.5.3 Finite Representations of Semi-Linear Sets 217
9.5.4 Complete Languages for FO + POLY-Expressible
Linear Queries 226

9.5.5	Complete Languages for Linear Queries	227
9.6	Bibliographic Notes	227
10.	Topological Queries	231
	Bart Kuijpers and Victor Vianu	
10.1	Introduction	231
10.2	Preliminaries	234
10.3	Languages for Topological Queries	236
10.3.1	The 4-Intersection Relations	236
10.3.2	Region-Based Languages	239
10.3.3	Topological Elementary Equivalence	249
10.4	Topological Invariants	257
10.4.1	Lossless Topological Invariants	257
10.4.2	Spatial Representations of Topological Information ...	261
10.5	Using Topological Invariants to Answer Topological Queries..	261
10.5.1	Fixpoint Queries on Topological Invariants	262
10.5.2	Translating Spatial Queries to Queries on the Invariant	265
10.6	Bibliographic Notes	271
11.	Euclidean Query Languages	275
	Bart Kuijpers, Gabriel Kuper, and Jan Paredaens	
11.1	Introduction	275
11.2	Semi-Circular Relations	277
11.3	Query Language over Encodings of Semi-Circular Relations ..	279
11.4	Safe Restriction of the Language	282
11.5	Languages for Semi-Circular Relations	283
11.5.1	Definition of the Query Language	284
11.5.2	Comparison with FO + LIN	286
11.5.3	Comparison with FO + POLY	288
11.6	Conclusion	290
11.7	Bibliographic Notes	290
12.	Genericity in Spatial Databases	293
	Bart Kuijpers and Dirk Van Gucht	
12.1	Introduction	293
12.2	Definitions and Examples	294
12.3	Undecidability Results	297
12.4	Sound and Complete Languages for Generic Queries	298
12.4.1	FO + POLY	298
12.4.2	Computable Queries	301
12.5	Bibliographical Notes	302

13. Linear Repeating Points	305
Pierre Wolper	
13.1 Introduction	305
13.2 A Constraint Model of Temporal Databases	306
13.2.1 Expressiveness of the Model	307
13.3 Computing with Temporal Constraints	308
13.3.1 A First Approach	308
13.3.2 Finite Automata as Constraints	309
13.3.3 Computing with Lrps Represented by Automata	312
13.4 Bibliographic Notes	313

Part III. Algorithmic Aspects

14. Optimization Techniques	319
Stéphane Grumbach, Zoé Lacroix, Philippe Rigaux, and Luc Segoufin	
14.1 Introduction	319
14.2 Spatial Query Processing and Optimization	320
14.2.1 Data Modeling	320
14.2.2 Query Languages	321
14.2.3 Query Processing	321
14.2.4 Query Optimization	322
14.3 Impact of the Data Format	323
14.3.1 Constraint Clustering	323
14.3.2 Orthographic Dimension	325
14.4 Query Processing	328
14.4.1 Alternation of Computation Modes	329
14.4.2 Query Pattern Recognition	330
14.5 Bibliographic Notes	333
15. Constraint Algebras	335
Dina Goldin	
15.1 Introduction	335
15.2 $FO(<)$	335
15.2.1 Data Representation	335
15.2.2 Canonical Form	336
15.3 Monotone Two-Variable Constraints	339
15.4 Bibliographic Notes	342

16. I/O-Efficient Algorithms for CDBs	343
Sridhar Ramaswamy	
16.1 Introduction	343
16.2 Dynamic Interval Management in Secondary Memory	345
16.2.1 Efficient Static Data Structure for Stabbing Queries ..	345
16.2.2 Dynamic Interval Management	348
16.2.3 Practical Aspects of Indexing Constraints	351
16.3 Constraint Join	352
16.3.1 One-Dimensional Join	352
16.3.2 Two-Dimensional Case: Rectangle Join	353
16.3.3 Practical Aspects of the Join	356
16.4 Lower Bounds	357
16.5 Conclusion	359
16.6 Bibliographic Notes	359

Part IV. Prototypes

17. The DEDALE Prototype	365
Stéphane Grumbach, Philippe Rigaux, Michel Scholl, and Luc Segoufin	
17.1 Introduction	365
17.2 The Data Model	366
17.3 Constraint Representation and Storage with O_2	368
17.3.1 Data Conversion: Loading and Displaying	369
17.4 The Query Language	371
17.5 Query Processing	375
17.5.1 Translation, Rewriting, and Evaluation	375
17.5.2 Constraint Manipulation	377
17.5.3 Implementation of Algebraic Operators	379
17.6 Conclusion	380
17.7 Bibliographic Notes	381
 18. The DISCO System	 383
Peter Revesz	
18.1 Introduction	383
18.2 DISCO Queries	383
18.3 Implementation	384
18.3.1 Converting to Relational Algebra	384
18.3.2 Optimization of Relational Algebra	386
18.4 Extensibility of DISCO	388
18.5 Bibliographic Notes	388

19. SQL/TP: A Temporal Extension of SQL	391
David Toman	
19.1 Introduction	391
19.2 Temporal Data Model and Constraint Encoding	391
19.2.1 Representable Temporal Databases	392
19.2.2 Data Definition Language	392
19.3 Queries	393
19.3.1 Syntax and Semantics	394
19.3.2 Query Compilation	395
19.4 Bibliographic Notes	398
 Bibliography	 401
 Index	 423

List of Contributors

Michael Benedikt

Bell Labs/Lucent Technologies
263 Shuman Blvd.
PO Box 3050
Naperville,
IL 60566, USA
benedikt@bell-labs.com

Jan Chomicki

Dept. of Computer Science
Monmouth University
Howard Hall
West Long Branch,
NJ 07764, USA
chomicki@monmouth.edu

Dina Goldin

Dept. of Math. and CS
UMASS-Boston
100 Morrissey Blvd.
Boston,
MA 02125, USA
dgg@cs.umb.edu

Stéphane Grumbach

INRIA
Rocquencourt BP105
78153 Le Chesnay Cedex
France
Stephane.Grumbach@inria.fr

Marc Gyssens

University of Limburg (LUC)
Dept. WNI
Universitaire Campus
3590 Diepenbeek
Belgium
gyssens@charlie.luc.ac.be

Bart Kuijpers

University of Limburg (LUC)
Dept. WNI
Universitaire Campus
3590 Diepenbeek
Belgium
bart.kuijpers@luc.ac.be

Zoé Lacroix

Data Logic
A Division of Gene Logic Inc.
2001 Center Street
Suite 600
Berkeley,
CA 94704, USA
zlacroix@genelogic.com

Sridhar Ramaswamy

Bell Labs/Lucent Technologies
600 Mountain Ave.
Murray Hill,
NJ 07974, USA
sridhar@bell-labs.com

XVI List of Contributors

Peter Revesz

Dept. of Comp. Sci. and Eng.
University of Nebraska-Lincoln
Lincoln,
NE 68588, USA
revesz@cse.unl.edu

Philippe Rigaux

Cedric/CNAM
292, rue St. Martin
75141 Paris Cedex 03
France
rigaux@cnam.fr

Michel Scholl

Cedric/CNAM
292, rue St. Martin
75141 Paris Cedex 03
France
and
INRIA
Rocquencourt BP105
78153 Le Chesnay Cedex
France
Michel.Scholl@inria.fr

Luc Segoufin

INRIA
Rocquencourt BP105
78153 Le Chesnay Cedex
France
Luc.Segoufin@inria.fr

Jianwen Su

Dept. of Computer Science
University of California
Santa Barbara,
CA 93106, USA
su@cs.ucsb.edu

David Toman

Dept. of Computer Science
200 University Avenue West
University of Waterloo
Waterloo,
Ontario N2L 3G1, Canada
david@algonquin.uwaterloo.ca

Jan Van den Bussche

University of Limburg (LUC)
Dept. WNI
Universitaire Campus
3590 Diepenbeek
Belgium
jbussche@cs.ucsd.edu

Luc Vandeurzen

University of Limburg (LUC)
Dept. WNI
Universitaire Campus
3590 Diepenbeek
Belgium
lvdeurze@luc.ac.be

Dirk Van Gucht

Computer Science Dept.
Indiana University
Bloomington,
IN 47405, USA
vgucht@cs.indiana.edu

Pierre Wolper

Institut Montefiore, B28
Université de Liège
4000 Liège
Belgium
pw@montefiore.ulg.ac.be

Victor Vianu

CSE Dept.
U. of California at San Diego
La Jolla,
CA 92093-0114, USA
vianu@cs.ucsd.edu