Bart Kuijpers
Peter Revesz (Eds.)

LNCS 3074

Constraint Databases

First International Symposium, CDB 2004
Paris, France, June 2004
Proceedings

Springer

Bart Kuijpers Peter Revesz (Eds.)

Constraint Databases

First International Symposium, CDB 2004
Paris, France, June 12-13, 2004
Proceedings

Springer

Volume Editors

Bart Kuijpers
University of Limburg (LUC)
Research Group on Theoretical Computer Science WNI
Universitaire Campus, 3590 Diepenbeek, Belgium
E-mail: bart.kuijpers@luc.ac.be

Peter Revesz
University of Nebraska-Lincoln
Computer Science and Engineering
214 Ferguson Hall, Lincoln, NE 68588-0115, USA
E-mail: revesz@cse.unl.edu

Library of Congress Control Number: 2004106087

CR Subject Classification (1998): H.2, H.3

ISSN 0302-9743
ISBN 3-540-22126-3 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11009566 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3074

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Preface

The first International Symposium on the Applications of Constraint Databases (CDB 2004) took place in Paris, France, on June 12–13, 2004, just before the ACM SIGMOD and PODS conferences.

Since the publication of the paper "Constraint Query Languages" by Kanellakis, Kuper and Revesz in 1990, the last decade has seen a growing interest in constraint database theory, query evaluation, and applications, reflected in a variety of conferences, journals, and books. Constraint databases have proven to be extremely flexible and adoptable in environments that relational database systems cannot serve well, such as geographic information systems and bioinformatics.

This symposium brought together people from several diverse areas all contributing to the practice and the application of constraint databases. It was a continuation and extension of previous workshops held in Friedrichshafen, Germany (1995), Cambridge, USA (1996), Delphi, Greece (1997), and Seattle, USA (1998) as well as of the work in the comprehensive volume "Constraint Databases" edited by G. Kuper, L. Libkin and J. Paredaens (2000) and the textbook "Introduction to Constraint Databases" by P. Revesz (2002).

The aim of the symposium was to open new and future directions in constraint database research; to address constraints over domains other than the reals; to contribute to a better implementation of constraint database systems, in particular of query evaluation; to address efficient quantifier elimination; and to describe applications of constraint databases.

The technical program of the symposium consisted of 10 technical papers and an invited paper as well as additional invited talks by Leonid Libkin and Andreas Podelski. The papers collected in these proceedings were selected by the program committee from a total of 29 submissions, and they were presented in five sessions, as described below.

Efficient query evaluation. Joos Heintz (invited speaker) and Bart Kuijpers address the difficulty of the effective evaluation of first-order queries, usually involving some form of quantifier elimination, and discuss various aspects that influence the efficiency of the evaluation of queries expressible in first-order logic over the reals. The importance of data structures and their effect on the complexity of quantifier-elimination is emphasized and a novel data model that supports data exploration and visualization as well as efficient query evaluation is proposed. Finally, they show that a particular kind of sample point query cannot be evaluated in polynomial time.

Spatial and spatio-temporal data. Spatial databases is a common application area of constraint databases. In recent years spatio-temporal data have often been modeled using constraints. We have three technical papers on this topic.

- Lixin Li, Youming Li and Reinhard Piltner propose a new spatio-temporal interpolation method for 3-D space and 1-D time geographic data, based on shape functions. Instead of only manipulating the time dimension as in the earlier ST product and tetrahedral methods, their new method takes the original approach of combining 2-D shape functions in the (x, y) domain with the (z, t) domain shape functions.
- Floris Geerts deals with the representation of moving objects in databases. Moving objects are usually represented, when possible, through explicit descriptions of their trajectories. The author proposes instead a new data model based on encoding their equations of motion, more specifically by differential equations. He also discusses a query language for this data model.
- Sofie Haesevoets describes a triangle-based logic in which queries that are invariant under affinities of the ambient space can be formulated. She characterizes the expressive power of this logic and shows it to be equivalent to the affine-generic fragment of first-order logic over the reals. She also presents algorithms for computing an affine-invariant triangulation and covering.

Applications. Looking at specific applications is important for two reasons. First, they reveal the possibilities of constraint database applications, often applications that could not be done in relational database systems. Second, they test the limits of the current constraint data model and query language proposals and thereby stimulate their further extensions. The following specific applications raise important issues and provide big challenges to researchers for the future.

- Maria Teresa Gómez López, Rafale Ceballos Guerrero, Rafael Martínez Gasca and Carmelo del Valle Sevilla apply constraint databases in the determination of potential minimal conflicts, which can be further used for polynomial model-based diagnosis.
- Viswanathan Ramanathan and Peter Revesz apply constraint databases to the genome map assembly problem. The genome map assembly problem is the problem of reconstructing the entire genome sequence of an organism based on overlapping fragments of the genome. They look at several algorithms for this problem. Using extensive computer experiments, they show that their constraint automaton, which can be solved using a constraint database system, solves the genome map assembly problem computationally more efficiently than the common alternative solution based on overlap multigraphs. Even more surprisingly, the average case running time of their solution increases only linearly while the running time of the other solution increases exponentially with the size of real genome data input.
- Carson Kai-Sang Leung proposes a new dynamic FP-Tree mining algorithm to mine frequent itemsets satisfying succinct constraints. The proposed algorithm is dynamic, such that the constraints can be changed during the mining process. Based on a classification of constraints this paper describes the cases of relaxing and tightening constraints and evaluation results show the effectiveness of this approach.

Query optimization. Query optimization is the concern of making the evaluation of queries computationally efficient in space and time. These techniques are essential elements for the implementation of constraint database systems. We had two papers in this area.

- Jan Chomicki discusses the problem of semantic query optimization for preference queries and treats this problem as a constraint reasoning problem. His techniques make use of integrity constraints, and make it possible to remove redundant occurrences of the winnow operator resulting in a more efficient algorithm for the computation of winnow. The paper also investigates the problem of propagating integrity constraints.
- Anagh Lal and Berthe Y. Choueiry consider the important problem of efficient join computation during query evaluation. They model the join computation in relational databases as a constraint satisfaction problem, which they solve using their technique called dynamic bundling. With dynamic bundling the join computation can be performed with major savings in space and time.

The future of constraint databases. Implementation of constraint databases is, of course, a major practical concern. While there are several prototype systems developed at universities and research laboratories, such as the C^3, the DEDALE and the MLPQ systems, there are still no commercial implementations of constraint databases. However, this situation may change in the future, as explained in the following two papers.

- Dina Goldin describes how constraints can be eliminated from constraint databases, in the sense of reducing them to as simple a representation as used in relational database systems and geographic information systems. She proposes a 3-tier architecture for constraint databases, with an abstract layer for the infinite relational extent of the data and a concrete layer that admits both constraint-based and geometry-based representations of spatiotemporal data.
- Mengchui Cai, from the DB2 group at the IBM Silicon Valley Laboratory, presents a way of integrating constraint databases into relational database systems. His main insight is that existing relational database systems can be extended by special functions that call a constraint relational engine at the appropriate places within an extended SQL query, while the constraint data itself can be represented within specialized relational tables. This proposal may lead to a practical and seemless way of integrating constraint data with relational data.

This symposium would have been impossible without the help and effort of many people. The editors would like to thank the program committee for the selection of the papers and the local organizers, in particular Irène Guessarian, for the arrangements in Paris. We especially would like to thank Sofie Haesevoets for managing the conference Web site and many other practical arrangements, and Floris Geerts for advertising the symposium and composing these proceedings.

The organizers are extremely grateful for the financial support given by General Eleftherios and Argyroula Kanellakis, the University of Limburg (LUC) and the University of Nebraska-Lincoln.

We would explicitly like to thank the Université Pierre et Marie Curie, Paris 6, for hosting the symposium.

We are pleased to bring to the reader these symposium proceedings, which reflect major recent advances in the field of constraint databases. We were also glad to see the symposium bring together many researchers in the field of constraint databases for a fruitful exchange of ideas. We also remembered those who due to their untimely death could not attend the symposium, including Paris Kanellakis and his family. Finally, we look forward to a continued growth in the field and to future symposium events.

June 2004 Bart Kuijpers and Peter Revesz

Conference Organization

Chairs

Bart Kuijpers	Limburgs Universitair Centrum, Belgium
Peter Revesz	University of Nebraska-Lincoln, USA

Program Committee

Saugata Basu	Georgia Tech, USA
Alex Brodsky	George Mason University, USA
Jan Chomicki	SUNY at Buffalo, USA
Berthe Choueiry	University of Nebraska-Lincoln, USA
Giorgio Delzanno	Università di Genova, Italy
Floris Geerts	University of Helsinki, Finland
Marc Giusti	CNRS, Ecole Polytechnique, Paris, France
Dina Goldin	University of Connecticut, USA
Stephane Grumbach	INRIA, Paris, France
Joxan Jaffar	National University of Singapore, Singapore
Manolis Koubarakis	Technical University of Crete, Greece
Stephan Kreutzer	Humboldt Universität, Berlin, Germany
Bart Kuijpers	Limburgs Universitair Centrum, Belgium
Gabriel Kuper	Università di Trento, Italy
Zoe Lacroix	Arizona State University, USA
Lixin Li	Georgia Southern University, USA
Jan Paredaens	Universiteit Antwerpen, Belgium
Peter Revesz	University of Nebraska-Lincoln, USA
Philippe Rigaux	Université Paris Sud, France
Kai-Uwe Sattler	Technische Universität Ilmenau, Germany
Jianwen Su	University of California at Santa Barbara, USA
David Toman	University of Waterloo, Canada
Jan Van den Bussche	Limburgs Universitair Centrum, Belgium
Dirk Van Gucht	Indiana University, USA
Nicolai Vorobjov	University of Bath, UK
Mark Wallace	Imperial College, London, UK

External Reviewers

Xiang Fu, Cagdas Gerede, Sofie Haesevoets, Anagh Lal, Hoda Mokhtar

Publicity and Proceedings Chair

Floris Geerts	University of Helsinki, Finland

Lecture Notes in Computer Science

For information about Vols. 1–2968

please contact your bookseller or Springer-Verlag

Vol. 3083: W. Emmerich, A.L. Wolf (Eds.), Component Deployment. X, 249 pages. 2004.

Vol. 3074: B. Kuijpers, P. Revesz (Eds.), Constraint Databases and Applications. XII, 181 pages. 2004.

Vol. 3065: A. Lomuscio, D. Nute (Eds.), Deontic Logic. X, 275 pages. 2004. (Subseries LNAI).

Vol. 3064: D. Bienstock, G. Nemhauser (Eds.), Integer Programming and Combinatorial Optimization. XI, 445 pages. 2004.

Vol. 3063: A. Llamosí, A. Strohmeier (Eds.), Reliable Software Technologies - Ada-Europe 2004. XIII, 333 pages. 2004.

Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), Advances in Artificial Intelligence. XIII, 582 pages. 2004. (Subseries LNAI).

Vol. 3059: C.C. Ribeiro, S.L. Martins (Eds.), Experimental and Efficient Algorithms. X, 586 pages. 2004.

Vol. 3058: N. Sebe, M.S. Lew, T.S. Huang (Eds.), Computer Vision in Human-Computer Interaction. X, 233 pages. 2004.

Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), Advances in Knowledge Discovery and Data Mining. XIX, 713 pages. 2004. (Subseries LNAI).

Vol. 3054: I. Crnkovic, J.A. Stafford, H.W. Schmidt, K. Wallnau (Eds.), Component-Based Software Engineering. XI, 311 pages. 2004.

Vol. 3053: C. Bussler, J. Davies, D. Fensel, R. Studer (Eds.), The Semantic Web: Research and Applications. XIII, 490 pages. 2004.

Vol. 3052: W. Zimmermann, B. Thalheim (Eds.), Abstract State Machines 2004. Advances in Theory and Practice. XII, 235 pages. 2004.

Vol. 3047: F. Oquendo, B. Warboys, R. Morrison (Eds.), Software Architecture. X, 279 pages. 2004.

Vol. 3046: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications - ICCSA 2004. LIII, 1016 pages. 2004.

Vol. 3045: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1040 pages. 2004.

Vol. 3044: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1140 pages. 2004.

Vol. 3043: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), Computational Science and Its Applications – ICCSA 2004. LIII, 1180 pages. 2004.

Vol. 3042: N. Mitrou, K. Kontovasilis, G.N. Rouskas, I. Iliadis, L. Merakos (Eds.), NETWORKING 2004, Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications. XXXIII, 1519 pages. 2004.

Vol. 3038: M. Bubak, G.D.v. Albada, P.M. Sloot, J. Dongarra (Eds.), Computational Science - ICCS 2004. XXI, 1118 pages. 2004.

Vol. 3037: M. Bubak, G.D.v. Albada, P.M. Sloot, J. Dongarra (Eds.), Computational Science - ICCS 2004. XXI, 776 pages. 2004.

Vol. 3036: M. Bubak, G.D.v. Albada, P.M. Sloot, J. Dongarra (Eds.), Computational Science - ICCS 2004. XXI, 776 pages. 2004.

Vol. 3035: M.A. Wimmer (Ed.), Knowledge Management in Electronic Government. XII, 326 pages. 2004. (Subseries LNAI).

Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), Advances in Web Intelligence. XIII, 227 pages. 2004. (Subseries LNAI).

Vol. 3033: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), Grid and Cooperative Computing. XXXVIII, 1076 pages. 2004.

Vol. 3032: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), Grid and Cooperative Computing. XXXVII, 1112 pages. 2004.

Vol. 3031: A. Butz, A. Krüger, P. Olivier (Eds.), Smart Graphics. X, 165 pages. 2004.

Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), Agent-Oriented Information Systems. XIV, 207 pages. 2004. (Subseries LNAI).

Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), Innovations in Applied Artificial Intelligence. XXI, 1272 pages. 2004.

Vol. 3028: D. Neuenschwander, Probabilistic and Statistical Methods in Cryptology. X, 158 pages. 2004.

Vol. 3027: C. Cachin, J. Camenisch (Eds.), Advances in Cryptology - EUROCRYPT 2004. XI, 628 pages. 2004.

Vol. 3026: C. Ramamoorthy, R. Lee, K.W. Lee (Eds.), Software Engineering Research and Applications. XV, 377 pages. 2004.

Vol. 3025: G.A. Vouros, T. Panayiotopoulos (Eds.), Methods and Applications of Artificial Intelligence. XV, 546 pages. 2004. (Subseries LNAI).

Vol. 3024: T. Pajdla, J. Matas (Eds.), Computer Vision - ECCV 2004. XXVIII, 621 pages. 2004.

Vol. 3023: T. Pajdla, J. Matas (Eds.), Computer Vision - ECCV 2004. XXVIII, 611 pages. 2004.

Vol. 3022: T. Pajdla, J. Matas (Eds.), Computer Vision - ECCV 2004. XXVIII, 621 pages. 2004.

Vol. 3021: T. Pajdla, J. Matas (Eds.), Computer Vision - ECCV 2004. XXVIII, 633 pages. 2004.

Vol. 3019: R. Wyrzykowski, J. Dongarra, M. Paprzycki, J. Wasniewski (Eds.), Parallel Processing and Applied Mathematics. XIX, 1174 pages. 2004.

Vol. 3016: C. Lengauer, D. Batory, C. Consel, M. Odersky (Eds.), Domain-Specific Program Generation. XII, 325 pages. 2004.

Vol. 3015: C. Barakat, I. Pratt (Eds.), Passive and Active Network Measurement. XI, 300 pages. 2004.

Vol. 3014: F. van der Linden (Ed.), Software Product-Family Engineering. IX, 486 pages. 2004.

Vol. 3012: K. Kurumatani, S.-H. Chen, A. Ohuchi (Eds.), Multi-Agnets for Mass User Support. X, 217 pages. 2004. (Subseries LNAI).

Vol. 3011: J.-C. Régin, M. Rueher (Eds.), Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. XI, 415 pages. 2004.

Vol. 3010: K.R. Apt, F. Fages, F. Rossi, P. Szeredi, J. Váncza (Eds.), Recent Advances in Constraints. VIII, 285 pages. 2004. (Subseries LNAI).

Vol. 3009: F. Bomarius, H. Iida (Eds.), Product Focused Software Process Improvement. XIV, 584 pages. 2004.

Vol. 3008: S. Heuel, Uncertain Projective Geometry. XVII, 205 pages. 2004.

Vol. 3007: J.X. Yu, X. Lin, H. Lu, Y. Zhang (Eds.), Advanced Web Technologies and Applications. XXII, 936 pages. 2004.

Vol. 3006: M. Matsui, R. Zuccherato (Eds.), Selected Areas in Cryptography. XI, 361 pages. 2004.

Vol. 3005: G.R. Raidl, S. Cagnoni, J. Branke, D.W. Corne, R. Drechsler, Y. Jin, C.G. Johnson, P. Machado, E. Marchiori, F. Rothlauf, G.D. Smith, G. Squillero (Eds.), Applications of Evolutionary Computing. XVII, 562 pages. 2004.

Vol. 3004: J. Gottlieb, G.R. Raidl (Eds.), Evolutionary Computation in Combinatorial Optimization. X, 241 pages. 2004.

Vol. 3003: M. Keijzer, U.-M. O'Reilly, S.M. Lucas, E. Costa, T. Soule (Eds.), Genetic Programming. XI, 410 pages. 2004.

Vol. 3002: D.L. Hicks (Ed.), Metainformatics. X, 213 pages. 2004.

Vol. 3001: A. Ferscha, F. Mattern (Eds.), Pervasive Computing. XVII, 358 pages. 2004.

Vol. 2999: E.A. Boiten, J. Derrick, G. Smith (Eds.), Integrated Formal Methods. XI, 541 pages. 2004.

Vol. 2998: Y. Kameyama, P.J. Stuckey (Eds.), Functional and Logic Programming. X, 307 pages. 2004.

Vol. 2997: S. McDonald, J. Tait (Eds.), Advances in Information Retrieval. XIII, 427 pages. 2004.

Vol. 2996: V. Diekert, M. Habib (Eds.), STACS 2004. XVI, 658 pages. 2004.

Vol. 2995: C. Jensen, S. Poslad, T. Dimitrakos (Eds.), Trust Management. XIII, 377 pages. 2004.

Vol. 2994: E. Rahm (Ed.), Data Integration in the Life Sciences. X, 221 pages. 2004. (Subseries LNBI).

Vol. 2993: R. Alur, G.J. Pappas (Eds.), Hybrid Systems: Computation and Control. XII, 674 pages. 2004.

Vol. 2992: E. Bertino, S. Christodoulakis, D. Plexousakis, V. Christophides, M. Koubarakis, K. Böhm, E. Ferrari (Eds.), Advances in Database Technology - EDBT 2004. XVIII, 877 pages. 2004.

Vol. 2991: R. Alt, A. Frommer, R.B. Kearfott, W. Luther (Eds.), Numerical Software with Result Verification. X, 315 pages. 2004.

Vol. 2990: J. Leite, A. Omicini, L. Sterling, P. Torroni (Eds.), Declarative Agent Languages and Techniques. XII, 281 pages. 2004. (Subseries LNAI).

Vol. 2989: S. Graf, L. Mounier (Eds.), Model Checking Software. X, 309 pages. 2004.

Vol. 2988: K. Jensen, A. Podelski (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. XIV, 608 pages. 2004.

Vol. 2987: I. Walukiewicz (Ed.), Foundations of Software Science and Computation Structures. XIII, 529 pages. 2004.

Vol. 2986: D. Schmidt (Ed.), Programming Languages and Systems. XII, 417 pages. 2004.

Vol. 2985: E. Duesterwald (Ed.), Compiler Construction. X, 313 pages. 2004.

Vol. 2984: M. Wermelinger, T. Margaria-Steffen (Eds.), Fundamental Approaches to Software Engineering. XII, 389 pages. 2004.

Vol. 2983: S. Istrail, M.S. Waterman, A. Clark (Eds.), Computational Methods for SNPs and Haplotype Inference. IX, 153 pages. 2004. (Subseries LNBI).

Vol. 2982: N. Wakamiya, M. Solarski, J. Sterbenz (Eds.), Active Networks. XI, 308 pages. 2004.

Vol. 2981: C. Müller-Schloer, T. Ungerer, B. Bauer (Eds.), Organic and Pervasive Computing – ARCS 2004. XI, 339 pages. 2004.

Vol. 2980: A. Blackwell, K. Marriott, A. Shimojima (Eds.), Diagrammatic Representation and Inference. XV, 448 pages. 2004. (Subseries LNAI).

Vol. 2979: I. Stoica, Stateless Core: A Scalable Approach for Quality of Service in the Internet. XVI, 219 pages. 2004.

Vol. 2978: R. Groz, R.M. Hierons (Eds.), Testing of Communicating Systems. XII, 225 pages. 2004.

Vol. 2977: G. Di Marzo Serugendo, A. Karageorgos, O.F. Rana, F. Zambonelli (Eds.), Engineering Self-Organising Systems. X, 299 pages. 2004. (Subseries LNAI).

Vol. 2976: M. Farach-Colton (Ed.), LATIN 2004: Theoretical Informatics. XV, 626 pages. 2004.

Vol. 2973: Y. Lee, J. Li, K.-Y. Whang, D. Lee (Eds.), Database Systems for Advanced Applications. XXIV, 925 pages. 2004.

Vol. 2972: R. Monroy, G. Arroyo-Figueroa, L.E. Sucar, H. Sossa (Eds.), MICAI 2004: Advances in Artificial Intelligence. XVII, 923 pages. 2004. (Subseries LNAI).

Vol. 2971: J.I. Lim, D.H. Lee (Eds.), Information Security and Cryptology -ICISC 2003. XI, 458 pages. 2004.

Vol. 2970: F. Fernández Rivera, M. Bubak, A. Gómez Tato, R. Doallo (Eds.), Grid Computing. XI, 328 pages. 2004.

Table of Contents

Efficient Query Evaluation

Spatial and Spatio-Temporal Data

Applications

Query Optimization

The Future of Constraint Databases

Constraint Databases, Data Structures and Efficient Query Evaluation*

Joos Heintz[1,2,3] and Bart Kuijpers[4]

[1] Universitad de Buenos Aires
Departamento de Computación
Ciudad Universitaria, Pabellón I, 1428 Buenos Aires, Argentina
joos@dc.uba.ar
[2] Consejo Nacional de Investigaciones
Científicas y Tecnológicas (CONICET), Argentina
[3] Universidad de Cantabria
Facultad de Ciencias
Departamento de Matemáticas, Estadística y Comptación
Avda. de los Castros s/n, E-39005 Santander, Spain
[4] Limburgs Universitair Centrum
Department WNI
Universitaire Campus, 3590 Diepenbeek, Belgium
bart.kuijpers@luc.ac.be

Abstract. Constraint databases that can be described by boolean combinations of polynomial inequalities over the reals have received ample research attention. In particular, the expressive power of first-order logic over the reals, as a constraint database query language, has been studied extensively. The difficulty of the effective evaluation of first-order queries, usually involving some form of quantifier elimination, has been largely neglected.

The contribution of this paper is a discussion of various aspects that influence the efficiency of the evaluation of queries expressible in first-order logic over the reals. We emphasize the importance of *data structures* and their effect on the complexity of quantifier-elimination. We also propose a novel data model that supports data exploration and visualization as well as efficient query evaluation. In this context, we introduce the concept of *sample point query*. Finally, we show that a particular kind of sample point query cannot be evaluated in polynomial sequential time by means of branching-parsimonious procedures.

1 Introduction and Summary

The framework of *constraint databases* was introduced in 1990 by Kanellakis, Kuper and Revesz [26] as an extension of the relational model that allows the

* Research partially supported by the following Argentinian, Belgian, German and Spanish grants: UBACyT X198, PIP CONICET 2461, FW/PA/02–EIII/007, ALA 01–E3/02 and DGCyT BFM 2000–0349.

B. Kuijpers and P. Revesz (Eds.): CDB 2004, LNCS 3074, pp. 1–24, 2004.

use of possibly infinite, but first-order definable relations rather than just finite relations. It provides an elegant and powerful model for applications that deal with infinite sets of points in some real affine space $\mathbb{R}^n$. In the setting of the constraint model, infinite relations are finitely represented as boolean combinations of polynomial equalities and inequalities, which we interpret, in this paper, over the real and exceptionally over the complex numbers. The case of the interpretation over the real numbers has applications in spatial databases [31].

Various aspects of the constraint model are well-studied by now (for an overview of research results we refer to [28] and the textbook [33]). The relational calculus augmented with polynomial constraints, or equivalently, first-order logic over the reals augmented with relation predicates to address the database relations $R_1, \ldots, R_s$, $\mathsf{FO}(+, \times, <, 0, 1, R_1, \ldots, R_s)$ for short, is the standard first-order query language for constraint databases. The expressive power of first-order logic over the reals, as a constraint database query language, has been studied extensively. However, the difficulty of the efficient evaluation of first-order queries, usually involving some form of quantifier elimination, has been largely neglected. The existing constraint database systems are based on general purpose quantifier-elimination algorithms and are, in most cases, restricted to work with linear data, i.e., they use first-order logic over the reals without multiplication [28, Part IV].

The intrinsic inefficiency of quantifier elimination represents a bottle-neck for real-world implementations of constraint database systems. General purpose elimination algorithms (such as, e.g., [6, 12, 18, 23, 32]) and standard data structures prevent query evaluation to become efficient. The fact that the knapsack problem can be formulated in this setting indicates that geometric elimination may be intrinsically hard. Another example for this complexity phenomenon is given by the permanent of a generic $n \times n$ matrix, which appears as the output of a suitable first-order query (see [22] for details on both examples).

In the literature of constraint databases, the data model proposed to describe geometric figures in $\mathbb{R}^n$ is based on quantifier-free first-order formulas over the reals [28, 33]. The data structures needed to implement this data model are left largely unspecified. It is widely understood that these formulas should be represented by explicitly giving disjunctive normal forms using dense or sparse encoding of polynomials. However, disjunctive normal forms may be unnecessarily large and the sparse representation of elimination polynomials may be very inefficient. For example, the sparse representation of the determinant of a generic $n \times n$ matrix contains $n!$ terms. On the other hand, the determinant can be represented by an $O(n^3)$ arithmetic boolean circuit (with divisions) which describes the Gaussian elimination algorithm. This suggests the use of alternative data structures for the representation of the classical data model of constraint databases. Indeed, the use of *arithmetic boolean circuits* as alternative data structure allows the design of a new generation of elimination algorithms which produce an exponential time complexity gain compared to the most efficient algorithms using traditional data structures (see [36] for a definition of the notion of arithmetic boolean circuit and [22] for this kind of complexity results). Nevertheless, in terms of the syntactical size of the input, the worst-case sequential time complexity of the elimination of a single existential quantifier block by means of the new algorithms remains still exponential. However, when we measure the input in a semantic (i.e., geometric) way, as is achieved, e.g., by the *system degree*, elimination of a single existential quantifier block becomes polynomial in this quantity (see e.g. [3, 4, 13, 15]). Unfortunately, this does not suffice for the design of algorithms able to evaluate purely existential queries in sequential time which is polynomial in the number of bounded variables. In fact, an non-polynomial lower bound for the underlying elimination problem can be deduced from the $P_{\mathbb{R}} \neq NP_{\mathbb{R}}$ conjecture in the Blum-Shub-Smale complexity model over the real numbers [8].

Another shortcoming of the classical data model for constraint databases is that it does not support data exploration and local visualization. Indeed, a quantifier-free formula in disjunctive normal form, describing the output of a query, allows the answering of, for instance, the membership question, but it is does not allow an easy exhibition of the output, by, e.g., the production of *sample points*, or, for low dimensions, a visualization of the output. To increase the tangibility of the output, we suggest considering a new type of query that produces sample points. Furthermore, it could be desirable to support an exploration of the neighborhood of such a sample point. This could be achieved by representing the output by a cell decomposition consisting of cells which are non-empty open subsets of smooth real varieties. In this way, starting from any sample point, its neighborhood within its cell may be explored. In this sense, we propose a novel data model for constraint databases, consisting of *smooth cells accompanied by sample points*. The known most efficient elimination procedures produce naturally such output representations, a typical example being CAD [12]. Therefore, when constraint database theory invokes quantifier elimination in query evaluation, it should also incorporate these features of the existing elimination algorithms.

In this context, we extend the concept of sample point query to queries that give rationally parameterized families of polynomial functions as output. Such queries will be called *extended sample point queries*. The main outcome of the paper is a proof that extended sample point queries, associated to first-order formulas containing a *fixed* number of quantifier alternations, cannot be evaluated in polynomial sequential time by so-called "branching-parsimonious algorithms". This lower bound result suggest that further research on the complexity of query evaluation in constraint database theory should be directed towards the identification of database and query classes that have a strongly improved complexity behavior. As a pattern for the development of such a theory, we may consider a new type of elimination algorithms which are based on the notion of system degree and use non-conventional data structures (see [2–4, 13, 15, 17, 20, 21, 24, 30, 34]).

This paper introduces a number of new concepts for constraint database theory that sometimes require certain notions from algebraic complexity theory, algebraic geometry and commutative algebra. These notions can be found in