Lecture Notes in Computer Science 3074

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

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

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

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

Table of Contents

Efficient Query Evaluation

Constraint Databases, Data Structures and Efficient Query Evaluation 1
Joos Heintz and Bart Kuijpers

Spatial and Spatio-Temporal Data

A New Shape Function Based Spatiotemporal Interpolation Method 25
Lixin Li, Youming Li, and Reinhard Piltner

Moving Objects and Their Equations of Motion 40
Floris Geerts

A Triangle-Based Logic for Affine-Invariant Querying
of Two-Dimensional Spatial Data 52
Sofie Haesevoets

Applications

Applying Constraint Databases in the Determination
of Potential Minimal Conflicts to Polynomial Model-Based Diagnosis 74
Maria Teresa Gómez López, Rafael Ceballos Guerrero, Rafael Martínez Gasca, and Carmelo del Valle Sevilla

Constraint Database Solutions to the Genome Map Assembly Problem ... 88
Viswanathan Ramanathan and Peter Revesz

Dynamic FP-Tree Based Mining
of Frequent Patterns Satisfying Succinct Constraints 112
Carson Kai-Sang Leung

Query Optimization

Semantic Optimization of Preference Queries 128
Jan Chomicki

Constraint Processing Techniques for Improving Join Computation:
A Proof of Concept 143
Anagh Lal and Berthe Y. Choueiry

The Future of Constraint Databases

Taking Constraints out of Constraint Databases 161
Dina Q. Goldin

Integrating Constraint and Relational Database Systems 173
Mengchu Cai

Author Index 181