# Report on the 1st International Symposium on the Applications of Constraint Databases (CDB'04)

Bart Kuijpers
Limburgs Universitair Centrum
Belgium
bart.kuijpers@luc.ac.be

Peter Revesz
University of Nebraska
USA
revesz@cse.unl.edu

## 1   Introduction

The 1st International Symposium on the Applications of Constraint Databases (CDB'04) was held on June 12–13, 2004, just before the ACM SIGMOD and PODS conferences, in the Amphithéatre de Chimie of the Université Pierre et Marie Curie in Paris, France. We acted as program committee chairs and Irène Guessarian and Patrick Cégielski as local organization chairs.

This symposium has brought together a group of around 30 researchers from diverse areas that contributed to both the the application and the theory 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).

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 in a variety of conferences, journals, and books. The symposium opened new directions in constraint database research by addressing constraints over domains other than the reals, by contributing to a better implementation of constraint database systems, in particular to query evaluation, by addressing efficient quantifier elimination, and by describing novel applications of constraint databases.

The technical program of the symposium consisted of 3 invited talks, 10 presentations of selected contributed papers and a lively and fruitful panel discussion. The invited talks were held by Leonid Libkin (University of Toronto, Canada), Joos Heintz (Universities of Buenos Aires, Argentina and Cantabria, Spain) and Andreas Podelski (Max-Planck-Institut für Informatik, Germany). The 10 contributing papers were selected by an international program committee of 26 researchers from a field of 28 submissions. The panel participants included Alex Brodsky (George Mason University, USA), Joos Heintz, Andreas Podelski, Jan Van den Bussche (Limburgs Universitair Centrum, Belgium) and Moshe Vardi (Rice University, USA). The symposium proceedings were published by Springer-Verlag as Volume 3074 of the Lecture Notes in Computer Science series.

## 2   Invited talks

The invited talks covered a variety of topics in constraint databases.

**Leonid Libkin**, in his talk "Constraints and Queries over Strings and Trees," discussed constraint databases over discrete domains, such as strings and trees, both ranked and unranked. While early research concentrated mostly on continuous domains (due to applications of constraint databases in querying geographical data), the focus has recently switched to discrete domains. Libkin gave a survey of recent results on decidable constraints over strings and trees that arise from automatic structures, and of query languages based on such constraints.

**Joos Heintz**, in his talk and paper "Constraint Databases, Data Structures and Efficient Query Evaluation," addressed the difficulty of the effective evaluation of first-order queries, usually involving some form of quantifier elimination and discussed 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 was emphasized and a novel data model that supports data exploration and visualization as well as efficient query evaluation was proposed. Finally, he showed that a particular kind of sample point query cannot be evaluated in polynomial time. This paper was joint work with Bart Kuijpers.

**Andreas Podelski**, in his talk "Constraint-Based Model Checking of ECA Rules," discussed the automatic verification of termination, confluence and other proper-

ties (safety and liveness) for Event Condition Action rules by combing recent approaches to constraint-based abstraction with query evaluation techniques for constraint databases. This talk was based in part on joint work with Hassan Ait-Kaci.

# 3 Contributed papers

The technical program was broken down into 4 sessions discussed here.

## 3.1 Spatial and spatio-temporal data

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

**Lixin Li**, **Youming Li** and **Reinhard Piltner**, in their paper "A New Shape Function Based Spatiotemporal Interpolation Method," 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**, in his paper "Moving objects and their equations of motion," 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**, in her paper "A triangle-based logic for affine-invariant querying of two-dimensional spatial data," 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.

## 3.2 Applications of constraint databases

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

**Maria Teresa Gómez López**, **Rafael Ceballos Guerrero**, **Rafael Martínez Gasca** and **Carmello del Valle Sevilla**, in their paper "Applying Constraint Databases in the Determination of Potential Minimal Conflicts to Polynomial Model-based Diagnosis", 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**, in their paper "Constraint Database Solutions to the Genome Map Assembly Problem," address the problem of reconstructing the entire genome sequence of an organism based on overlapping fragments of its 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, is much more efficient in solving the genome map assembly problem than is 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, in the paper "Dynamic FP-Tree Based Mining of Frequent Patterns Satisfying Succinct Constraints," a new dynamic FP-Tree mining algorithm to mine frequent itemsets satisfying succinct constraints. The proposed algorithm is dynamic, that is, the constraints can be changed during the mining process. Based on a classification of constraints he describes the cases of relaxing and tightening constraints and extensive evaluation results showing the effectiveness of the new approach.

## 3.3 Query optimization

Query optimization is concerned with making computationally efficient, in space and time, the evaluation of queries. Good query optimization techniques are essential for the implementation of constraint database systems. The symposium had two papers in this area.

**Jan Chomicki**, in the paper "Semantic Optimization of Preference Queries," 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 in their paper "Improving Join Computation Using Constraint

Processing Techniques" 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.

## 3.4 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 in "Taking Constraints out of Constraint Databases" 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 spatio-temporal data.

**Mengchui Cai**, from the DB2 group at the IBM Silicon Valley Laboratory, presents in the paper "Integrating constraint and relational database systems" 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 seamless way of integrating constraint data with relation data.

# 4  Panel Discussion

The symposium was concluded by a panel on the future of constraint databases. The panelists were Alex Brodsky, Joos Heintz, Andreas Podelski, Jan Van den Bussche, and Moshe Vardi, and it was moderated by Peter Revesz.

**Moshe Vardi** opened the discussion by asking *how theory can be turned into practice*. To illustrate this point he looked at model checking, which is a very successful field with numerous applications that started from an underlying theory of automaton on infinite words and resulted in many algorithms and good heuristics with good complexity. On the other hand, the field of dependencies, which

has no clear business problems to solve, studies certain fragments of first-order logic and resulted in many nice papers but remained without implementations and practice. Where is the area of constraint databases going? Will it turn out like the former or the later example? It is important to ask what are its applications, what problems it solves best, and how well compared to other possible solutions. For example, there is a possibility to develop heuristics that may lead to an efficient geographical information system based on constraint database theory.

**Jan Van den Bussche** mentioned that theory is still important to work on, and that it is impossible to see ahead whether a good theory will or will not have applications. As far as applications, he believes that bioinformatics could be a very fruitful application area of constraint databases.

**Andreas Podelski** emphasized that model checking in fact has a close relationship with constraint databases. Model checking could be solved by a combination of approximations and constraint database query evaluations. Hence he sees a potential for future growth in combining model checking with constraint databases.

**Alex Brodsky** described his company's experiences in software development that included in parts constraint processing with ideas based on constraint database techniques. He believes that more involvement of constraint databases in software development would be also possible with further improvement and commercialization of constraint database systems.

**Joos Heintz** was the last panelist to express his views. He also emphasized the need for combining theory and practice, but he also pointed to the need to be patient. He comes from mathematics to the area of constraint databases, and he believes that compared to many areas of mathematics, constraint databases looks much more promising for future applications and success.

There was a very active participation in the panel discussion by numerous members of the audience. Overall the panel was a lively and stimulating event.

# 5  Summary

The symposium had a very active program and participation. We were glad to see the symposium bring together many researchers in the field of constraint databases for a fruitful exchange of ideas. We also look forward to a continued growth in the field and to future events in the field of constraint databases.