Introduction to Constraint Databases

by Peter Revesz

published in 2002 by Springer in the series *Texts in Computer Science*,
407 pages, 112 illustrations, hardcover, list price $54.95
ISBN 0-387-98729-0

Review by:
Bart Kuijpers, University of Limburg (Belgium)
brt.kuijpers@luc.ac.be

Context and aim of the book

This book is the first textbook on constraint databases. Its author, together with P. Kanellakis and G. Kuper, introduced constraint databases in 1990 as a powerful generalization of the relational database model. Constraints, such as linear or polynomial equalities and inequalities, are used to finitely represent possibly infinite sets of points. They provide an elegant way to combine classical relational data with, for instance, spatial or temporal data.

Since the early 1990s, the topic of constraint databases has received considerable research interest, both theoretical and towards systems development, and has been present at most database conferences during the past decade. It turned out to be a rich area in which a combination of techniques from, e.g., logic, (finite) model theory, algebraic and computational geometry, topology, query languages and symbolic computation are applied. A comprehensive survey of the research results in this field appeared two years ago (*Constraint databases*, edited by G. Kuper, L. Libkin and J. Paredaens, Springer, 2000). Whereas this survey mainly addresses researchers, the present book aims at making the topic of constraint databases accessible to advanced undergraduate and beginning graduate students and mainly addresses constraint databases from a developer's point of view. This book will certainly contribute to the exposure of constraint databases to a wider audience and hopefully also to its proliferation in a broader database practice.

Summary of the book

This textbook presents the constraint database model as a powerful extension of the relational model that allows a user or programmer to work easily with infinite data. It covers a wide range of constraint formalisms and shows that the constraint model provides an elegant tool for data modeling and querying in application areas such as geographic information systems (GIS), spatiotemporal data management, bioinformatics, genome databases and computer vision. This book covers a substantial part of constraint-database theory, emphasizes several developer's issues, and presents a number of sample constraint database systems. The author starts by developing the constraint data model from the relational one. Next, he shows how familiar query languages, such as the relational algebra, SQL, and various forms of Datalog carry over to the constraint model. A third broad part of the book focuses on query evaluation and addresses theoretical topics such as quantifier-elimination algorithms for several constraint languages and the complexity of query evaluation in these languages. Also more specific data models and query languages are addressed, e.g., for spatiotemporal database applications. Next, the book describes a sample linear constraint database system, a Boolean constraint database system, and a spatiotemporal database system. A last part of the book presents a number of sample applications.

A walk through the chapters

The first two chapters of the book present a smooth development of the relational data model into the constraint data model. Also on the level of query languages the approach of developing constraint languages from traditional relational languages is followed. As a basic set of query languages, the relational algebra, SQL and various forms of Datalog with or without aggregation and negation are presented in the next three chapters. Chapters 6 and 7 offer an intermezzo into constraint automata and refinement queries.

The core part of the book is devoted to query evaluation and presents constraint databases from a developer's point of view (Chapters 8-16). Here, a first chapter is devoted to the problem of safety (the developer faces the well-know problem that recursion involving arithmetic over an infinite domain is not guaranteed to terminate). Various techniques to guarantee safety are discussed. Tools to test whether a query is safe and safety certification by expert programmers are presented further on (Chapter 11). Chapters 9 and 10 offer more theoretical background. Quantifier-elimination algorithms are presented for several constraint formalisms. Next, the evaluation of relational algebra, SQL and Datalog queries is described and an analysis of the complexity of query evaluation is presented. A number of low-level details for the implementation of quantifier elimination and constraint database systems in general are addressed. In Chapter 12, two implementation approaches are described in detail (to the level of data structures and pseudo-code). Other topics that are covered are interoperability among different types of constraint databases (Chapter 14), approximate representation of data by constraints (Chapter 15) and static data visualization (Chapter 16). There are a number of chapters that specifically address problems relevant to spatiotemporal databases (Chapter 13). They cover dynamic data visualization, data storage and indexing issues (Chapter 17), and the PReSTO system that works with parametric rectangles (Chapter 20).

In the last quarter of the book, a number of systems and sample applications are presented. MLPQ is a system for linear constraints that supports Datalog queries (Chapter 18), and DISCO implements Datalog with Boolean constraints (Chapter 19). The sample constraint database applications in the last three chapters deal with computer vision, bioinfomatics (also genome data), and environmental modeling.

Target audience

This book aims at advanced undergraduate and beginning graduate students. As indicated and illustrated in the preface, the book can be used for a variety of introductory and advanced courses. The book is largely self-contained and provides the necessary background. Elementary knowledge of linear algebra and programming languages is assumed, and for some parts, a basic background in databases would be recommendable. To facilitate learning, at the end of each chapter extensive exercise sets are provided and sample software systems of real-world applications are available on the author's web page. But the book also addresses database experts and designers. The chapters contain valuable bibliographical references for further reading.

Reviewer's appreciation

Up to now the rich field of constraint databases was only accessible through research papers and the earlier mentioned research survey. This book takes a unique place and will certainly play an important role in spreading the ideas behind constraint databases among a wider audience, especially among students. Hopefully it will eventually also contribute to a broader constraint database practice. There is certainly a need for a book like this.

Although the book devotes much attention to topics from its author's research, it offers a strong background in constraint-database theory and comprehensively addresses developer's issues. I would certainly recommend this book for teaching constraint databases, as I would recommend teaching this topic in advanced database courses in general.