

Edited by G. Goos and J. Hartmanis

Advisory Board: W. Brauer D. Gries J. Stoer



G. Grahne

The Problem of Incomplete Information in Relational Databases

Springer-Verlag

Berlin Heidelberg New York

London Paris Tokyo

Hong Kong Barcelona

Budapest

Series Editors

Gerhard Goos
Universität Karlsruhe
Postfach 69 80
Vincenz-Priessnitz-Straße 1
W-7500 Karlsruhe, FRG

Juris Hartmanis
Department of Computer Science
Cornell University
5148 Upson Hall
Ithaca, NY 14853, USA

Author

Gösta Grahne
University of Helsinki, Department of Computer Science
Teollisuuskatu 23, SF-00510 Helsinki, Finland

CR Subject Classification (1991): H.2.1-2, F.2.2, I.2.3-4

ISBN 3-540-54919-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-54919-6 Springer-Verlag New York Berlin Heidelberg

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, re-use of illustrations, recitation, broadcasting, reproduction on microfilms 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 1991
Printed in Germany

Typesetting: Camera ready by author
Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
45/3140-543210 - Printed on acid-free paper

Preface

In practice it is often the case that the available information is incomplete with respect to the information that is supposed to be recorded in a database. This monograph considers the problems raised by information incompleteness in the context of the relational model.

The basic semantic assumption is that an incomplete database represents a set of complete databases (relations). We show that there are two natural lattice structures on the set of all sets of relations. These lattices enable us to give precise meanings to operations performed on incomplete databases. The operations are querying, dependency enforcement and updates.

There are several candidate tools for storing and manipulating databases with incomplete information. We focus on generalizations of relations. These generalizations, called *tables*, allow variables representing unknown values as entries, as well as entries constraining the variables. We define four increasingly (syntactically) strict classes of tables and we study their abilities to serve as a basis for implementing incomplete databases and the operations on them. It turns out that the new class of *Horn tables* is the most significant from a practical point of view. In this class of tables we are able to efficiently evaluate positive existential queries and in some cases least fixpoint (recursive) queries. In Horn tables we can also efficiently incorporate the information contained in one join dependency and a set of equality generating dependencies, as well as hold the result of a subclass of the update operations.

An earlier version of the present work was submitted as the author's Ph. D. dissertation in 1989 to the Department of Computer Science at the University of Helsinki. With respect to the process of writing the thesis, I want to make the following acknowledgements.

The best way to learn a craft is to work under the guidance of a skilled professional. Professor Kari-Jouko R  ih   kindly initiated me in this way into computer science. He has given me invaluable support through his constructive advice and through his balance and wisdom. I am proud of being his student.

I was very fortunate in coming into contact with Professor Esko Ukkonen. His deep insight and sharp technical advice, as well as his warm-hearted encouragement, have been a great source of inspiration for me. May the sun always shine for him.

I would like to express my appreciation of Professor Martti Tienari, who has created an excellent research milieu, with fine facilities and a high standard for the staff to live up to. It has been both a pleasure and a challenge to work at his department.

The research community has many distinguished members, some of whom I would like to mention in addition, since they have willingly shared their time and thoughts with me. They are Serge Abiteboul, Paris Kanellakis, the late Witold Lipski, Pekka Orponen, Michalis Spyrtos, and Nicolas Spyrtos. I am especially grateful to Serge and Paris, who granted me the privilege of collaborating with them.

Finally, I would like to thank all my friends and colleagues at the Department of Computer Science for creating a stimulating atmosphere to work in. The financial support of the Academy of Finland and the Ministry of Education is also gratefully acknowledged. In addition, thanks are also due to the referees of Springer-Verlag, whose reports I have used in trying to improve the presentation of the material, and to the Department of Computer Science at the University of Toronto, who provided the necessary logistics for the present work.

Toronto
September 1991

G. Grahne

Contents

1. Introduction	1
2. Relational databases	6
2.1 The relational model and query languages	6
2.2 Data dependencies	15
2.3 Updates to databases	25
2.4 Computational complexity	30
3. Semantic aspects of incomplete information	33
3.1 Incomplete databases and queries	33
3.2 Dependencies and incomplete information	44
3.3 Updating incomplete databases	46
4. Syntactic and algorithmic aspects of incomplete information	55
4.1 Tables as representations of incomplete information	55
4.2 Querying tables	61
4.3 Dependencies and tables	77
4.4 Updates and tables	92
5. Computational complexity aspects of incomplete information	104
5.1 The complexity of query evaluation	104
5.2 The complexity of dependency enforcement	122
5.3 The complexity of updates	130
6. Some conclusive aspects	132
6.1 Relation to other work	132
6.2 A conclusion	140

References	143
Index of notation	150
Appendix. Problems used in reduction proofs	154
A1. 3CNF Satisfiability	154
A2. 3DNF Tautology	155
A3. HC Satisfiability	156
A4. Bipartite graph matching	156