Quantifiers in Action

Generalized Quantification in Query, Logical and Natural Languages

ADVANCES IN DATABASE SYSTEMS
Volume 37

Series Editors

Ahmed K. Elmagarmid
Purdue University
West Lafayette, IN 47907

Amit P. Sheth
Wright State University
Dayton, Ohio 45435

For other titles published in this series, go to
www.springer.com/series/5573

Quantifiers in Action

Generalized Quantification in Query, Logical and Natural Languages

by

Antonio Badia
Computer Engineering and Computer Science Department
Speed School of Engineering
University of Louisville
Louisville, KY, USA

Springer

Author:
Antonio Badia
Computer Engineering and Computer Science Department
JB. Speed School of Engineering
University of Louisville
Louisville, KY 40292 USA
abadia@louisville.edu

Library of Congress Control Number: 2009921815

ISBN: 978-0-387-09563-9 e-ISBN: 978-0-387-09564-6

© Springer Science+Business Media, LLC 2009
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Mindy, Emma and Gabi: I'm finally done!

Preface

The goal of this book is to help conquer what this author sees as dangerous divides in the study of databases in general, and of query languages in particular: the divide between theory and practice, and the divide between purely formal approaches and those with a wider concept of communication.

This is, of course, a very tall order, and there is no claim here that the mission has been accomplished. We live in an era of specialists (people who know a lot about a little), and the explosion of available material in any given field makes it quite difficult even to keep current with one's area, never mind exploring a different one. Presenting all the material this book touches upon in depth is nearly impossible. Rather, the book introduces the most basic results of different fields. It is then likely that an expert in any given area (logic, databases, or formal linguistics) will find that the material discussed here is very elementary. This is part of the price to pay when trying to reach the widest audience. The book is intended just to present the beginning of an exploration, and its aim is to encourage others to pursue research further. Necessarily, the coverage is limited and subjective, but hopefully it will become an *entry point* for the interested reader, who will be encouraged to delve further into the subject using some of the bibliography provided.

It may seem strange to pick a a technical, specialized topic like Generalized Quantification for this kind of endeavor. However, Generalized Quantifiers are a perfect example of a practical theory. An idea that started in the seminal papers by Mostowski ([76]) and Lindstrom ([70]) as a theoretical investigation into a new concept found unsuspected applications first in linguistics and then in query languages. Generalized Quantifiers motivate insights into optimization of set-based queries, Question Answering, and Cooperative Query Answering. This book tries to gather all those insights together. As such, the book tries to walk a line between theory and practice. There is out there a large collection of work on GQs; however, they all tend towards the purely theoretical side ([4]) or towards a particular application, usually linguistics ([94]). The book strives the cover the middle ground.

Depending on their point of view, readers may feel that we have gone too far to the theory side or to the application side. Others still may feel that we have not gone far enough. There are no doubt many other possible applications of the concept, and much more is known right now that we can hope to include in our first chapters. However, by bringing together all this material we hope to make people aware of what others are doing. If we happen to encourage collaborations and help establish links where before there were none, then we consider that we have done our job.

Before getting into the subject, it is time to give credit where credit is due. This book represents a summary of research carried out by the author over the past years. In particular, section 4.3 of chapter 4 is based on [10]; sections 6.2, 6.3 and 6.4 of chapter 6 are based on [13]; chapter 7 is based on [9] and chapter 8 is an update of work reported in [11]. Along the way, there were many people who helped carry out the research, and this is the opportunity to thank them, as they made this book possible. This list should certainly include the members of my doctoral dissertation committee (Dirk van Gucht, Ed Robertson, David Leake and Larry Moss), as it all got started there. The collaboration with Stijn Vansummeren was a pleasure; I learned much from it. My graduate students contributed considerably to this effort; special mention should go to Bin Cao, Brandon Debes and Dazhuo Li for their work in different parts of the project. Bin helped develop the interpreter introduced in chapter 5; Dazhuo worked an approach to linear prefixes, shown in chapter 6; and Brandon implemented a system that transformed our project from a paper-and-pencil approach to executable software[1]. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0347555, and therefore special mention goes to the National Science Foundation, which generously supported the research described here under a CAREER Award, and to my two program managers, Le Gruenwald and Maria Zemankova, who exhibited a tremendous amount of patience with me -I am very indebted to both of them.

Finally, I cannot honestly say that Mindy, Emma or Gabi's contributions have been technical. In fact, I cannot say that repeatedly asking "Tickle me, papa!" is a contribution at all. But I know that without them, I would have never spent the time and energy I did on this, or any other project -it would just not seem worthwhile.

Louisville, KY

Antonio Badia
December 2008

[1] The result of his efforts is accessible at `http://qlgq.spd.louisville.edu/index.php`.

Contents

1 Introduction 1

2 Basic Concepts 7
- 2.1 From Propositional to First Order Logic 7
- 2.2 Quantification 8
  - 2.2.1 Semantics 9
  - 2.2.2 Meaning 11
- 2.3 More on Quantification 12
  - 2.3.1 Quantifier Scope and Prefixes 12
  - 2.3.2 Skolemization 14
  - 2.3.3 Quantifier Rank 15
  - 2.3.4 Relativization 17
- 2.4 Games 18
- 2.5 More Semantics 19
  - 2.5.1 Expressive Power of FOL 22
  - 2.5.2 Finite and Infinite Models 22

3 Generalized Quantifiers 25
- 3.1 Introduction 25
- 3.2 Generalized Quantifiers 25
- 3.3 Another view 30
- 3.4 Basic Complexity 33

4 *QLGQ*: A Query Language with Generalized Quantifiers . . . 37
- 4.1 Introduction: GQs in query languages 37
- 4.2 $QLGQ$ 38
  - 4.2.1 Syntax of $QLGQ$ 39
  - 4.2.2 Semantics of $QLGQ$ 41
  - 4.2.3 Remarks on Syntax 42
- 4.3 Safety and Domain Independence 44
  - 4.3.1 Relation to other languages 49

4.4 Generalized Quantifiers and SQL 50

5 Implementation and Optimization of Standard GQs 55
5.1 Languages to Define GQs 55
5.2 Translating and Optimizing *QLGQ* 60
5.3 The Interpreter 60
5.3.1 Complex Queries 66
5.4 Optimization 67
5.4.1 Optimization on RA Expressions 67
5.4.2 Optimization using GQ Properties 67
5.5 Application to SQL 69
5.6 Monadic vs. Polyadic Quantification 71

6 Quantifier Prefixes 73
6.1 Introduction 73
6.1.1 Linear and Non-linear Prefixes 74
6.1.2 Henkin Prefixes and Generalized Quantifiers 76
6.2 Linear and Non-Linear Prefixes in *QLGQ* 77
6.3 Cumulation 81
6.4 Branching 82
6.5 Linear Prefixes 85
6.5.1 Algebraic Translation 87

7 Cooperative Query Answering 91
7.1 Introduction 91
7.2 Cooperative Query Answering 91
7.3 Cooperative Query Answering with QLGQ 94
7.3.1 Presuppositions 94
7.3.2 Constructing Explanations and Justifications 97
7.3.3 Relaxed Queries 100
7.3.4 Expressing and Using Constraints 103
7.4 Further Research in CQA 105

8 Generalized Quantifiers and Natural Language 107
8.1 Introduction 107
8.2 Question Answering 107
8.3 GQs in Natural Language Analysis 110
8.3.1 Combining Quantifiers 114
8.4 QLGQ in QA 116
8.5 CQA, QA and GQs 120
8.6 Challenges 122

9 Extensions 127
9.1 Datalog-like Languages 127
9.1.1 Aggregates 127
9.1.2 Fixpoint 128
9.1.3 Higher Order Variables 131
9.2 Distributed Quantification 134
9.2.1 Quantification and Distributed Databases 135
9.2.2 Computing Distributed Quantification 139
9.3 Other Data Models 143

10 Conclusion 149

References 151