Texts in Theoretical Computer Science
An EATCS Series

Springer-Verlag Berlin Heidelberg GmbH

Leonid Libkin

Elements of Finite Model Theory

With 24 Figures

Springer

Author

Prof. Leonid Libkin
Department of Computer Science
University of Toronto
Toronto ON M5S 3H5
Canada
libkin@cs.toronto.edu
www.cs.toronto.edu/~libkin

Series Editors

Prof. Dr. Wilfried Brauer
Institut für Informatik der TUM
Boltzmannstr. 3, 85748 Garching, Germany
Brauer@informatik.tu-muenchen.de

Prof. Dr. Grzegorz Rozenberg
Leiden Institute of Advanced Computer Science
University of Leiden
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
rozenber@liacs.nl

Prof. Dr. Arto Salomaa
Turku Centre for Computer Science
Lemminkäisenkatu 14 A, 20520 Turku, Finland
asalomaa@utu.fi

Library of Congress Control Number: 2004105855

ACM Computing Classification (1998): F.4, F.1, H.2, F.3

ISBN 978-3-642-05948-3 ISBN 978-3-662-07003-1 (eBook)
DOI 10.1007/978-3-662-07003-1

springeronline.com

Originally published by Springer-Verlag Berlin Heidelberg New York in 2004
Softcover reprint of the hardcover 1st edition 2004

Cover Design: KünkelLopka, Heidelberg
Typesetting: Camera-ready by the author
Printed on acid-free paper 45/3142/GF - 5 4 3 2 1 0

To Helen and Daniel
Алёне и Даниле

Preface

Finite model theory is an area of mathematical logic that grew out of computer science applications.

The main sources of motivational examples for finite model theory are found in database theory, computational complexity, and formal languages, although in recent years connections with other areas, such as formal methods and verification, and artificial intelligence, have been discovered.

The birth of finite model theory is often identified with Trakhtenbrot's result from 1950 stating that validity over finite models is not recursively enumerable; in other words, completeness fails over finite models. The technique of the proof, based on encoding Turing machine computations as finite structures, was reused by Fagin almost a quarter century later to prove his celebrated result that put the equality sign between the class NP and existential second-order logic, thereby providing a machine-independent characterization of an important complexity class. In 1982, Immerman and Vardi showed that over ordered structures, a fixed point extension of first-order logic captures the complexity class PTIME of polynomial time computable properties. Shortly thereafter, logical characterizations of other important complexity classes were obtained. This line of work is often referred to as descriptive complexity.

A different line of finite model theory research is associated with the development of relational databases. By the late 1970s, the relational database model had replaced others, and all the basic query languages for it were essentially first-order predicate calculus or its minor extensions. In 1974, Fagin showed that first-order logic cannot express the transitive closure query over finite relations. In 1979, Aho and Ullman rediscovered this result and brought it to the attention of the computer science community. Following this, Chandra and Harel proposed a fixed-point extension of first-order logic on finite relational structures as a query language capable of expressing queries such as the transitive closure. Logics over finite models have become the standard starting point for developing database query languages, and finite model theory techniques are used for proving results about their expressiveness and complexity.

Yet another line of work on logics over finite models originated with Büchi's work from the early 1960s: he showed that regular languages are precisely those definable in monadic second-order logic over strings. This line of work is the automata-theoretic counterpart of descriptive complexity: instead of logical characterizations of time/space restrictions of Turing machines, one provides such characterizations for weaker devices, such as automata. More recently, connections between database query languages and automata have been explored too, as the field of databases started moving away from relations to more complex data models.

In general, finite model theory studies the behavior of logics on finite structures. The reason this is a separate subject, and not a tiny chapter in classical model theory, is that most standard model-theoretic tools (most notably, compactness) fail over finite models. Over the past 25–30 years, many tools have been developed to study logics over finite structures, and these tools helped answer many questions about complexity theory, databases, formal languages, etc.

This book is an introduction to finite model theory, geared towards theoretical computer scientists. It grew out of my finite model theory course, taught to computer science graduate students at the University of Toronto. While teaching that course, I realized that there is no single source that covers all the main areas of finite model theory, and yet is suitable for computer science students. There are a number of excellent books on the subject. *Finite Model Theory* by Ebbinghaus and Flum was the first standard reference and heavily influenced the development of the field, but it is a book written for mathematicians, not computer scientists. There is also a nice set of notes by Väänänen, available on the web. Immerman's *Descriptive Complexity* deals extensively with complexity-theoretic aspects of finite model theory, but does not address other applications. *Foundations of Databases* by Abiteboul, Hull, and Vianu covers many database applications, and Thomas's chapter "Languages, automata, and logic" in the *Handbook of Formal Languages* describes connections between logic and formal languages. Given the absence of a single source for all the subjects, I decided to write course notes, which eventually became this book.

The reader is assumed to have only the most basic computer science and logic background: some discrete mathematics, theory of computation, complexity, propositional and predicate logic. The book also includes a background chapter, covering logic, computability theory, and computational complexity. In general, the book should be accessible to senior undergraduate students in computer science.

A note on exercises: there are three kinds of these. Some are the usual exercises that the reader should be able to do easily after reading each chapter. If I indicate that an exercise comes from a paper, it means that its level could range from moderately to extremely difficult: depending on the exact level, such an "exercise" could be a question on a take-home exam, or even a course project, whose main goal is to understand the paper where the result is proven. Such exercises also gave me the opportunity to mention a number of interesting results that otherwise could not have been included in the book. There are also exercises marked with an asterisk: for these, I do not know solutions.

It gives me the great pleasure to thank my colleagues and students for their help. I received many comments from Marcelo Arenas, Pablo Barceló, Michael Benedikt, Ari Brodsky, Anuj Dawar, Ron Fagin, Arthur Fischer, Lauri Hella, Christoph Koch, Janos Makowsky, Frank Neven, Juha Nurmonen, Ben Rossman, Luc Segoufin, Thomas Schwentick, Jan Van den Bussche, Victor Vianu, and Igor Walukiewicz. Ron Fagin, as well as Yuri Gurevich, Alexander Livchak, Michael Taitslin, and Vladimir Sazonov, were also very helpful with historical comments. I taught two courses based on this book, and students in both classes provided very useful feedback; in addition to those I already thanked, I would like to acknowledge Antonina Kolokolova, Shiva Nejati, Ken Pu, Joseph Rideout, Mehrdad Sabetzadeh, Ramona Truta, and Zheng Zhang. Despite their great effort, mistakes undoubtedly remain in the book; if you find one, please let me know. My email is `libkin@cs.toronto.edu`.

Many people in the finite model theory community influenced my view of the field; it is impossible to thank them all, but I want to mention Scott Weinstein, from whom I learned finite model theory, and immediately became fascinated with the subject.

Finally, I thank Ingeborg Mayer, Alfred Hofmann, and Frank Holzwarth at Springer-Verlag for editorial assistance, and Denis Thérien for providing ideal conditions for the final proofreading of the book.

This book is dedicated to my wife, Helen, and my son, Daniel. Daniel was born one week after I finished teaching a finite model theory course in Toronto, and after several sleepless nights I decided that perhaps writing a book is the type of activity that goes well with the lack of sleep. By the time I was writing Chap. 6, Daniel had started sleeping through the night, but at that point it was too late to turn back. And without Helen's help and support I certainly would not have finished this book in only two years.

Toronto, Ontario, Canada
May 2004

Leonid Libkin

Contents