Answering Queries Using Views

Second Edition

Synthesis Lectures on Data Management

Synthesis Lectures on Data Management is edited by H.V. Jagadish of the University of Michigan. The series publishes 80–150 page publications on topics pertaining to data management. Topics include query languages, database system architectures, transaction management, data warehousing, XML and databases, data stream systems, wide scale data distribution, multimedia data management, data mining, and related subjects.

Answering Queries Using Views, Second Edition
Foto Afrati and Rada Chirkova
2019

Data Exploration Using Example-Based Methods
Matteo Lissandrini, Davide Mottin, Themis Palpanas, and Yannis Velegrakis
2018

Data Profiling
Ziawasch Abedjan, Lukasz Golab, Felix Naumann, and Thorsten Papenbrock
2018

Querying Graphs
Angela Bonifati, George Fletcher, Hannes Voigt, and Nikolay Yakovets
2018

Query Processing over Incomplete Databases
Yunjun Gao and Xiaoye Miao
2018

Natural Language Data Management and Interfaces
Yunyao Li and Davood Rafiei
2018

Human Interaction with Graphs: A Visual Querying Perspective
Sourav S. Bhowmick, Byron Choi, and Chengkai Li
2018

On Uncertain Graphs
Arijit Khan, Yuan Ye, and Lei Chen
2018

Answering Queries Using Views
Foto Afrati and Rada Chirkova
2017

Databases on Modern Hardware: How to Stop Underutilization and Love Multicores
Anatasia Ailamaki, Erieta Liarou, Pınar Tözün, Danica Porobic, and Iraklis Psaroudakis
2017

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, Media Restore, and System Failover, Second Edition
Goetz Graefe, Wey Guy, and Caetano Sauer
2016

Generating Plans from Proofs: The Interpolation-based Approach to Query Reformulation
Michael Benedikt, Julien Leblay, Balder ten Cate, and Efthymia Tsamoura
2016

Veracity of Data: From Truth Discovery Computation Algorithms to Models of Misinformation Dynamics
Laure Berti-Équille and Javier Borge-Holthoefer
2015

Datalog and Logic Databases
Sergio Greco and Cristina Molinaro
2015

Big Data Integration
Xin Luna Dong and Divesh Srivastava
2015

Instant Recovery with Write-Ahead Logging: Page Repair, System Restart, and Media Restore
Goetz Graefe, Wey Guy, and Caetano Sauer
2014

Similarity Joins in Relational Database Systems
Nikolaus Augsten and Michael H. Böhlen
2013

Information and Influence Propagation in Social Networks
Wei Chen, Laks V.S. Lakshmanan, and Carlos Castillo
2013

Data Cleaning: A Practical Perspective
Venkatesh Ganti and Anish Das Sarma
2013

Data Processing on FPGAs
Jens Teubner and Louis Woods
2013

Perspectives on Business Intelligence
Raymond T. Ng, Patricia C. Arocena, Denilson Barbosa, Giuseppe Carenini, Luiz Gomes, Jr., Stephan Jou, Rock Anthony Leung, Evangelos Milios, Renée J. Miller, John Mylopoulos, Rachel A. Pottinger, Frank Tompa, and Eric Yu
2013

Semantics Empowered Web 3.0: Managing Enterprise, Social, Sensor, and Cloud-based Data and Services for Advanced Applications
Amit Sheth and Krishnaprasad Thirunarayan
2012

Data Management in the Cloud: Challenges and Opportunities
Divyakant Agrawal, Sudipto Das, and Amr El Abbadi
2012

Query Processing over Uncertain Databases
Lei Chen and Xiang Lian
2012

Foundations of Data Quality Management
Wenfei Fan and Floris Geerts
2012

Incomplete Data and Data Dependencies in Relational Databases
Sergio Greco, Cristian Molinaro, and Francesca Spezzano
2012

Business Processes: A Database Perspective
Daniel Deutch and Tova Milo
2012

Data Protection from Insider Threats
Elisa Bertino
2012

Deep Web Query Interface Understanding and Integration
Eduard C. Dragut, Weiyi Meng, and Clement T. Yu
2012

P2P Techniques for Decentralized Applications
Esther Pacitti, Reza Akbarinia, and Manal El-Dick
2012

Query Answer Authentication
HweeHwa Pang and Kian-Lee Tan
2012

Declarative Networking
Boon Thau Loo and Wenchao Zhou
2012

Full-Text (Substring) Indexes in External Memory
Marina Barsky, Ulrike Stege, and Alex Thomo
2011

Spatial Data Management
Nikos Mamoulis
2011

Database Repairing and Consistent Query Answering
Leopoldo Bertossi
2011

Managing Event Information: Modeling, Retrieval, and Applications
Amarnath Gupta and Ramesh Jain
2011

Fundamentals of Physical Design and Query Compilation
David Toman and Grant Weddell
2011

Methods for Mining and Summarizing Text Conversations
Giuseppe Carenini, Gabriel Murray, and Raymond Ng
2011

Probabilistic Databases
Dan Suciu, Dan Olteanu, Christopher Ré, and Christoph Koch
2011

Peer-to-Peer Data Management
Karl Aberer
2011

Probabilistic Ranking Techniques in Relational Databases
Ihab F. Ilyas and Mohamed A. Soliman
2011

Uncertain Schema Matching
Avigdor Gal
2011

Fundamentals of Object Databases: Object-Oriented and Object-Relational Design
Suzanne W. Dietrich and Susan D. Urban
2010

Advanced Metasearch Engine Technology
Weiyi Meng and Clement T. Yu
2010

Web Page Recommendation Models: Theory and Algorithms
Sule Gündüz-Ögüdücü
2010

Multidimensional Databases and Data Warehousing
Christian S. Jensen, Torben Bach Pedersen, and Christian Thomsen
2010

Database Replication
Bettina Kemme, Ricardo Jimenez-Peris, and Marta Patino-Martinez
2010

Relational and XML Data Exchange
Marcelo Arenas, Pablo Barcelo, Leonid Libkin, and Filip Murlak
2010

User-Centered Data Management
Tiziana Catarci, Alan Dix, Stephen Kimani, and Giuseppe Santucci
2010

Data Stream Management
Lukasz Golab and M. Tamer Özsu
2010

Access Control in Data Management Systems
Elena Ferrari
2010

An Introduction to Duplicate Detection
Felix Naumann and Melanie Herschel
2010

Privacy-Preserving Data Publishing: An Overview
Raymond Chi-Wing Wong and Ada Wai-Chee Fu
2010

Keyword Search in Databases
Jeffrey Xu Yu, Lu Qin, and Lijun Chang
2009

Answering Queries Using Views, Second Edition
Foto Afrati and Rada Chirkova

ISBN: 978-3-031-00743-9	paperback
ISBN: 978-3-031-01871-8	ebook
ISBN: 978-3-031-00098-0	hardcover

DOI 10.1007/978-3-031-01871-8

A Publication in the Springer series
SYNTHESIS LECTURES ON DATA MANAGEMENT

Lecture #54
Series Editor: H.V. Jagadish, *University of Michigan*
Founding Editor: M. Tamer Özsu, *University of Waterloo*
Series ISSN
Print 2153-5418 Electronic 2153-5426

Answering Queries Using Views

Second Edition

Foto Afrati
National Technical University of Athens

Rada Chirkova
North Carolina State University

SYNTHESIS LECTURES ON DATA MANAGEMENT #54

ABSTRACT

The topic of using views to answer queries has been popular for a few decades now, as it cuts across domains such as query optimization, information integration, data warehousing, website design and, recently, database-as-a-service and data placement in cloud systems.

This book assembles foundational work on answering queries using views in a self-contained manner, with an effort to choose material that constitutes the backbone of the research. It presents efficient algorithms and covers the following problems: query containment; rewriting queries using views in various logical languages; equivalent rewritings and maximally contained rewritings; and computing certain answers in the data-integration and data-exchange settings. Query languages that are considered are fragments of SQL, in particular select-project-join queries, also called conjunctive queries (with or without arithmetic comparisons or negation), and aggregate SQL queries.

This second edition includes two new chapters that refer to tree-like data and respective query languages. Chapter 8 presents the data model for XML documents and the XPath query language, and Chapter 9 provides a theoretical presentation of tree-like data model and query language where the tuples of a relation share a tree-structured schema for that relation and the query language is a dialect of SQL with evaluation techniques appropriately modified to fit the richer schema.

KEYWORDS

conjunctive queries, aggregate queries, arithmetic comparisons, negation, acyclic queries, query containment, query equivalence, equivalent query rewritings, maximally contained query rewritings, tuple-generating dependencies, equality-generating dependencies, the chase algorithm query containment and rewriting under dependencies, data exchange, determinacy, XPath, tree data

Contents

Preface to the First Edition

Views are used in various scenarios; some of the view-based settings have been considered in depth by the research community. The settings that we cover in this book include the following.

1. The problem of rewriting queries using views, where a set of views and a set of queries are given, and we need to find equivalent rewritings (if they exist) of the queries using these views.

2. Sometimes we cannot find equivalent rewritings but can still compute a significant part of the answer to the query. This gives rise to the problems of computing certain answers and of finding maximally contained query rewritings.

3. The picture of finding rewritings changes when we assume that the data satisfy certain constraints (dependencies). We re-examine the problem of finding rewritings for the setting in which the constraints are tuple-generating dependencies and equality-generating dependencies.

4. A closely related topic based on the same theoretical foundations is the data-exchange setting. We define the concept of certain answers and present algorithms to find them.

5. Some theoretical aspects of the more general problem of answering queries using views have also been investigated in more general and abstract settings, such as determining the (query) language fragments for which there are rewritings in the cases where the queries are determined by the views.

In order to solve problems in the above settings, we need technical tools. Thus, we provide detailed treatments of some tools, including conjunctive-query containment (with and without arithmetic comparisons and negation), the chase algorithm for reasoning about dependencies, and going beyond nonrecursive languages to find rewritings, with a discussion of Datalog.

What the book is not about: This book is not about indexes or data structures that implement the techniques considered in the exposition. Instead, this books focuses on the formal-logic perspective on the topic.

The book is written in a linear way, in the sense that each chapter depends on all the previous chapters. We have made every effort for the book to be self-contained, thus there are no substantial prerequisites. A reader familiar with the basics of the theory of database systems and knowledge of logic will move faster through the chapters. Exercises are included. More exercises, including online exercises, and any supplementary material will be found on the website[1] of the book.

Foto Afrati and Rada Chirkova
October 2017

[1] bit.ly/2zMYBHf

Preface to the Second Edition

We have added two chapters that refer to tree-like data and respective query languages. Chapter 8 presents the data model for XML documents and the XPath query language and gives query containment tests and query rewriting techniques. However, besides the relational data model and the tree-structured XML data model, many recently developed systems, to manage big data, use data models that combine relational and tree-like features. In terms of defining and using views, these models have not be investigated. However, they are mature enough to be presented in a rigorous way. Thus, Chapter 9 provides a theoretical presentation of such data model and query language as extensions of the relational model and SQL query language. In these data models, the tuples of a relation share a tree-structured schema for that relation and the query language is a dialect of SQL. The query language uses SQL-style syntax but the evaluation techniques have to be modified to fit the richer schema. The conclusions and bibliographical notes for these two chapters are included in the chapters themselves. Chapters 8 and 9 can also be read independently of the rest of the chapters in the book and independently of each other.

In the chapters of the first edition, we have corrected errors.

Foto Afrati and Rada Chirkova
January 2019

Acknowledgments

Rachel Pottinger kindly reviewed this book and provided constructive comments. Jeff Ullman contributed immensely in an elegant and concise way in the improvement of this book. We acknowledge here his many comments that have greatly improved and enhanced the text.

Foto Afrati and Rada Chirkova
October 2017