Studies in Big Data

Volume 31

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Studies in Big Data" (SBD) publishes new developments and advances in the various areas of Big Data- quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence incl. neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/11970

Sergio Flesca · Sergio Greco
Elio Masciari · Domenico Saccà
Editors

A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years

Springer

Editors
Sergio Flesca
University of Calabria
Rende
Italy

Elio Masciari
ICAR-CNR
Rende
Italy

Sergio Greco
University of Calabria
Rende
Italy

Domenico Saccà
University of Calabria
Rende
Italy

ISSN 2197-6503 ISSN 2197-6511 (electronic)
Studies in Big Data
ISBN 978-3-319-61892-0 ISBN 978-3-319-61893-7 (eBook)
DOI 10.1007/978-3-319-61893-7

Library of Congress Control Number: 2017943818

© Springer International Publishing AG 2018
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Database research played a crucial role for the growth of the computer science field since its early stage. The research on this topic has greatly evolved during the last 25 years, and many research directions have been explored that had a relevant impact on several industrial applications: data mining and knowledge discovery, information extraction, semi-structured data and Web information systems to cite a few can be considered today as mature research fields with numerous algorithms and studies to manage heterogeneous data. Indeed, all these application scenarios have their roots in database field. Moreover, the advances in information technology lead to a new data management paradigm, referred to as Big Data, that will drive the research and the industry in the future.

The Italian Database Community has been active in all those research fields, and many well-reputed Italian scientists provided interesting original contributions.

This book sets out to give the reader a comprehensive view of the evolution of the database research field from classical relational database management systems to the current Big Data metaphor. The individual contributions of this book will illustrate the point of view of several research groups in Italy that had in the last 25 year quite active collaboration with leading research groups spread all around the world.

This book will be of great interest for readers willing to catch the evolution of the research on databases from a different perspective yet in an unified proposal. It is composed of five parts and a total of twenty-nine chapters.

Part I focuses on **Big Data** by illustrating several aspects ranging from data pre-elaboration to data warehousing. It consists of seven chapters. Part II analyses issues posed by **Incomplete/Inconsistent Data and Uncertain Reasoning** dealing with ontologies, schema mappings and uncertain queries. It consists of five chapters. Part III illustrates results related to **Data Modelling and Querying**. It consists of four chapters. Part IV analyses issues posed by **Knowledge Discovery and Data Mining** from relational data, social networks and complex data. It consists of eight chapters. Finally, Part V deals with **Security, Privacy and Health Systems** issues. It consists of five chapters.

We would like to thank all the authors who submitted their contributions for publication in this book. We also thank Thomas Ditzinger and Janusz Kacprzyk of Springer for their continuous support.

Rende, Italy
June 2017

Sergio Flesca
Sergio Greco
Elio Masciari
Domenico Saccà

Contents

Part I Big Data

The Visual Side of the Data 3
Marco Angelini, Tiziana Catarci, Massimo Mecella and Giuseppe Santucci

Digital Libraries: From Digital Resources to Challenges in Scientific Data Sharing and Re-Use 27
Maristella Agosti, Nicola Ferro and Gianmaria Silvello

From Data Integration to Big Data Integration 43
Sonia Bergamaschi, Domenico Beneventano, Federica Mandreoli, Riccardo Martoglia, Francesco Guerra, Mirko Orsini, Laura Po, Maurizio Vincini, Giovanni Simonini, Song Zhu, Luca Gagliardelli and Luca Magnotta

Matching Techniques for Data Integration and Exploration: From Databases to Big Data 61
Silvana Castano, Alfio Ferrara and Stefano Montanelli

25+ Years of Query Processing - From a Single, Stored Data Set to Big Data (and Beyond) 77
Barbara Catania and Giovanna Guerrini

From Star Schemas to Big Data: 20+ Years of Data Warehouse Research 93
M. Golfarelli and S. Rizzi

A Short Account of Techniques for Assisting Users in Mastering Big Data 109
Davide Martinenghi, Elisa Quintarelli, Fabio A. Schreiber and Letizia Tanca

Part II Incomplete/Inconsistent Data and Uncertain Reasoning

Multimedia, Similarity, and Preferences: Adding Flexibility to Your Information Needs 127
Ilaria Bartolini, Paolo Ciaccia and Marco Patella

Dealing with Inconsistency in Databases: An Overview 143
Marco Calautti, Luciano Caroprese, Bettina Fazzinga, Sergio Flesca, Filippo Furfaro, Sergio Greco, Cristian Molinaro, Francesco Parisi, Andrea Pugliese, Domenico Saccà, Irina Trubitsyna and Ester Zumpano

First-Order Ontology Mediated Database Querying via Query Reformulation 169
Diego Calvanese and Enrico Franconi

Using Ontologies for Semantic Data Integration 187
Giuseppe De Giacomo, Domenico Lembo, Maurizio Lenzerini, Antonella Poggi and Riccardo Rosati

Schema Mappings: From Data Translation to Data Cleaning 203
Giansalvatore Mecca, Paolo Papotti and Donatello Santoro

Part III Data Modeling and Querying

Data Modeling Across the Evolution of Database Technology 221
Paolo Atzeni, Luca Cabibbo and Riccardo Torlone

Modeling, Modeling, Modeling: From Web to Enterprise to Crowd to Social 235
Marco Brambilla and Stefano Ceri

Structural Decomposition Methods: Key Notions and Database Applications 253
G. Greco, N. Leone, F. Scarcello and G. Terracina

Multimedia Data Modeling and Management 269
Vincenzo Moscato and Antonio Picariello

Par IV Knowledge Discovery and Data Mining

How Data Mining and Machine Learning Evolved from Relational Data Base to Data Science 287
G. Amato, L. Candela, D. Castelli, A. Esuli, F. Falchi, C. Gennaro, F. Giannotti, A. Monreale, M. Nanni, P. Pagano, L. Pappalardo, D. Pedreschi, F. Pratesi, F. Rabitti, S. Rinzivillo, G. Rossetti, S. Ruggieri, F. Sebastiani and M. Tesconi

A Tour from Regularities to Exceptions 307
Fabrizio Angiulli, Fabio Fassetti, Luigi Palopoli and Domenico Ursino

Relational Data Mining in the Era of Big Data 323
Annalisa Appice, Michelangelo Ceci and Donato Malerba

Data Mining in Databases: Languages and Indices 341
Elena Baralis, Tania Cerquitelli, Silvia Chiusano and Rosa Meo

20+ Years of Analytics on Complex Data: Impact, Issues, Challenges and Contributions 353
Stefano Basta, Giuseppe Manco, Elio Masciari and Luigi Pontieri

Services Discovery and Recommendation for Multi-datasource Access: Exploiting Semantic and Social Technologies 375
Devis Bianchini, Valeria De Antonellis and Michele Melchiori

Data Semantics Meets Knowledge Discovery in Databases 391
Claudia Diamantini, Domenico Potena and Emanuele Storti

Analysing Trajectories of Mobile Users: From Data Warehouses to Recommender Systems 407
Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, Alessandra Raffaetà, Chiara Renso and Claudio Silvestri

Part V Security, Privacy and Health Systems

Big Data Security and Privacy 425
Elisa Bertino and Elena Ferrari

Not Only Databases: Social Data and Cybersecurity Perspective 441
Francesco Buccafurri, Gianluca Lax, Serena Nicolazzo and Antonino Nocera

Confidentiality Protection in Large Databases 457
Sabrina De Capitani di Vimercati, Sara Foresti, Giovanni Livraga, Stefano Paraboschi and Pierangela Samarati

Database Community and Health Related Data: Experiences Through the Last Decade 473
Pietro H. Guzzi, Giuseppe Tradigo and Pierangelo Veltri

From Rome to Milan, from Appennines to Alps, from Data to Big Data 489
Carlo Batini

Author Index 501