Lecture Notes in Computer Science 10494

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Edmon Begoli · Fusheng Wang
Gang Luo (Eds.)

Data Management and Analytics for Medicine and Healthcare

Third International Workshop, DMAH 2017
Held at VLDB 2017
Munich, Germany, September 1, 2017
Proceedings

Springer

Editors
Edmon Begoli
Oak Ridge National Laboratory
Oak Ridge, TN
USA

Gang Luo
University of Washington
Seattle, WA
USA

Fusheng Wang
Stony Brook University
Stony Brook, NY
USA

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-67185-7 ISBN 978-3-319-67186-4 (eBook)
DOI 10.1007/978-3-319-67186-4

Library of Congress Control Number: 2017952570

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

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

In this volume we present the accepted contributions for the Third International Workshop on Data Management and Analytics for Medicine and Healthcare (DMAH 2017), held in Munich, Germany, in conjunction with the 43rd International Conference on Very Large Data Bases (VLDB), on September 1st, 2017.

The goal of the workshop was to bring together researchers from the cross-cutting domains of research including information management and biomedical informatics. The workshop aimed to foster the exchange of information and discussions on innovative data management and analytics technologies. We encouraged topics that highlighted the end-to-end applications, systems and methods addressing problems in healthcare, public health, and everyday wellness; integration with clinical, physiological, imaging, behavioral, environmental, and "omics" data, as well as the data from social media and the Web. Our hope for this workshop was to provide a unique opportunity for mutual benefits and informative interaction between information management and biomedical researchers from the interdisciplinary fields.

A total of 16 papers were submitted to the DMAH workshop. A rigorous, single-blind peer-review selection mechanism was adapted, resulting in 9 accepted papers presented at the workshop. Each paper was reviewed by three members of the Program Committee, who were carefully selected for their knowledge and competence. As far as possible, papers were matched with the reviewers' particular interests and special expertise. The result of this careful process can be seen here in the high quality of the contributions published within this volume.

We would like to express our sincere thanks especially to the internationally renowned speakers who gave keynote talks at the workshop plenary sessions: Li Xiong from Emory University, USA, and Vagelis Hristidis from University of California, Riverside, USA.

We would like to thank the members of the Program Committee for their attentiveness, perseverance, and willingness to provide high-quality reviews.

July 2017

Edmon Begoli
Fusheng Wang
Gang Luo

Organization

Program Committee

Jesús B. Alonso-Hernández	Universidad de Las Palmas de Gran Canaria, Spain
Edmon Begoli	Oak Ridge National Laboratory, USA
Thomas Brettin	Argonne National Laboratory, USA
J. Blair Christian	Oak Ridge National Laboratory, USA
Carlo Combi	Università degli Studi di Verona, Italy
Kerstin Denecke	Bern University of Applied Sciences, Switzerland
Dejing Dou	University of Oregon, USA
Alevtina Dubovitskaya	EPFL, HES-SO, Switzerland
Peter Elkin	University at Buffalo, USA
Vijay Gadapally	MIT Lincoln Labs/CSAIL, USA
Zhe He	Florida State University, USA
Guoqian Jiang	Mayo Clinic College of Medicine, USA
Jun Kong	Emory University, USA
Tahsin Kurc	Stony Brook University, USA
Ulf Leser	Humboldt-Universität zu Berlin, Germany
Yanhui Liang	Stony Brook University, USA
Gang Luo	University of Washington, USA
Fernando Martin-Sanchez	Weill Cornell Medicine, USA
Wolfgang Mueller	Heidelberg Institute for Theoretical Studies, Germany
Casey Overby	Johns Hopkins University, USA
Fusheng Wang	Stony Brook University, USA
Hua Xu	The University of Texas School of Biomedical Informatics at Houston, USA

Abstracts of the Keynotes

Health Data Management and Analytics with Privacy and Confidentiality

Li Xiong

Emory University, Atlanta GA 30322, USA
lxiong@emory.edu

Abstract. Managing and analyzing large-scale clinical and public health data while protecting privacy of human subjects has been a key challenge in biomedical research. Traditional de-identification approaches are subject to various re-identification and disclosure risks and do not provide sufficient privacy protection for patients. This talk gives an overview of our work on privacy preserving health data sharing and analytics along two dimensions: (1) data encryption techniques that support secure computation and query processing on the encrypted data without disclosing the raw data, (2) differential privacy techniques that ensure the computation and query results do not disclose patient information. Focusing on the second dimension, a set of differential privacy techniques are presented that handle different types of data including relational, sequential, and time series data. Case studies using real health datasets are presented to demonstrate the feasibility of the solutions while outlining their limitations and open challenges.

Keywords: Health data management · Data analytics · Differential privacy · Data encryption · Secure computation

Acknowledgement. The presented work was supported by the National Institute of Health (NIH) under award number R01GM114612, R01GM118609-01A1, the Patient-Centered Outcomes Research Institute (PCORI) under contract ME-1310-07058.

Analysis of Online Health-Related User-Generated Content

Vagelis Hristidis

University of California, Riverside
vagelis@cs.ucr.edu

Abstract. An increasing amount of health-related content is posted online by patients, ranging from health forums to provider reviews. Analyzing this mostly text information can discover health trends and help patients make more informed decisions. We will discuss about the technical challenges involved in analyzing such data, including the use of biomedical ontologies, concept extraction, training set expansion and summarization. Completed and ongoing work on various application will be presented, as well as opportunities for future research.

Contents