Communications in Computer and Information Science 822

Commenced Publication in 2007
Founding and Former Series Editors:
Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu, Dominik Ślęzak, and Xiaokang Yang

Editorial Board

Simone Diniz Junqueira Barbosa
Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil

Joaquim Filipe
Polytechnic Institute of Setúbal, Setúbal, Portugal

Igor Kotenko
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia

Krishna M. Sivalingam
Indian Institute of Technology Madras, Chennai, India

Takashi Washio
Osaka University, Osaka, Japan

Junsong Yuan
University at Buffalo, The State University of New York, Buffalo, USA

Lizhu Zhou
Tsinghua University, Beijing, China

More information about this series at http://www.springer.com/series/7899

Leonid Kalinichenko · Yannis Manolopoulos
Oleg Malkov · Nikolay Skvortsov
Sergey Stupnikov · Vladimir Sukhomlin (Eds.)

Data Analytics and Management in Data Intensive Domains

XIX International Conference, DAMDID/RCDL 2017
Moscow, Russia, October 10–13, 2017
Revised Selected Papers

Springer

Editors
Leonid Kalinichenko
Federal Research Center
"Computer Science and Control"
Russian Academy of Sciences
Moscow
Russia

Yannis Manolopoulos
Open University of Cyprus
Latsia
Cyprus

Oleg Malkov
Institute of Astronomy
Russian Academy of Sciences
Moscow
Russia

Nikolay Skvortsov
Federal Research Center
"Computer Science and Control"
Russian Academy of Sciences
Moscow
Russia

Sergey Stupnikov
Federal Research Center
"Computer Science and Control"
Russian Academy of Sciences
Moscow
Russia

Vladimir Sukhomlin
Moscow State University
Moscow
Russia

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-319-96552-9 ISBN 978-3-319-96553-6 (eBook)
https://doi.org/10.1007/978-3-319-96553-6

Library of Congress Control Number: 2018948633

© Springer International Publishing AG, part of Springer Nature 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.

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

Preface

This CCIS volume published by Springer contains the proceedings of the XIX International Conference Data Analytics and Management in Data-Intensive Domains (DAMDID/RCDL 2017) that took place during October 9–13 in the Lomonosov Moscow State University at the Department of Computational Mathematics and Cybernetics. The DAMDID series of conferences was planned as a multidisciplinary forum of researchers and practitioners from various domains of science and research, promoting cooperation and exchange of ideas in the area of data analysis and management in domains driven by data-intensive research. Approaches to data analysis and management being developed in specific data-intensive domains (DID) of X informatics (such as X = astro, bio, chemo, geo, med, neuro, physics, chemistry, material science etc.), social sciences, as well as in various branches of informatics, industry, new technologies, finance, and business contribute to the conference content.

Traditionally DAMDID/RCDL proceedings are published locally before the conference as a collection of full texts of all regular and short papers accepted by the Program Committee as well as, abstracts of posters and demos. Soon after the conference, the texts of regular papers presented at the conference are submitted for online publishing in a volume of the European repository of the CEUR Workshop Proceedings, as well as for indexing the volume content in DBLP and Scopus. Since 2016, a DAMDID/RCDL volume of post-conference proceedings with up to one third of the submitted papers that were previously published in CEUR Workshop Proceedings have been published by Springer in their *Communications in Computer and Information Science* (CCIS) series. Each paper selected for the CCIS post-conference volume should be modified as follows: the title of each paper should be a new one; the paper should be significantly extended (with at least 30% new content); the paper should refer to its original version in the CEUR Workshop Proceedings. CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, and Scopus.

The program of DAMDID/RCDL 2017, as with the previous editions of these conferences, alongside the traditional data management topics reflects a rapid move into the direction of data science and data-intensive analytics. The program this year included carefully selected invited keynote talks related to rapidly developed DID. The respective plenary sessions were also aimed at attracting the attention of researchers in the selected DID. A preconference plenary session on October 9 included two talks: the keynote talk by Stefano Ceri, Professor of Database Systems at Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB) of Politecnico di Milano, and the invited talk by Zoltan Szallasi, MD, senior research scientist, the Children's Hospital Informatics Program, Harvard Medical School. The session was devoted to the development of methods and techniques for genomes and diagnostics in various application domains (from health care to criminalistics). Stefano Ceri considered the implementation issues of the new-generation DNA sequencing techniques in the

European project GeCo applying big data technologies; in the talk by Zoltan Szallasi, an overview of approaches to the genomic-based diagnostics in various application domains was given. In more detail, in the tutorial given by Zoltan Szallasi on October 10 the application of genomic diagnostics in cancer immunotherapy was presented. The problems of data deluge in astronomy and approaches to their solution were considered in the keynote talk by Giuseppe Longo (Professor of Astrophysics at the University of Naples Federico II). On the basis of their talks, Zoltan Szallasi, Stefano Ceri with co-authors, and Giuseppe Longo with co-authors provided invited full papers for this CCIS volume.

The conference Program Committee reviewed 75 submissions for the conference and eight submissions for the PhD workshop. For the workshop, five papers were accepted and three were rejected. For the conference, 47 submissions were accepted as full papers, 12 as short papers, two as posters, and two as demos, whereas 12 submissions were rejected. According to the conference program, these 59 oral presentations (of the full and short papers) are structured into 19 sessions including: Data Analysis Projects in Astronomy; Semantic Web Techniques in DID; Special Purpose DID Infrastructures (two sessions); Distributed Computing; System Efficiency Evaluation; Data Analysis Projects in Neuroscience; Specific Data Analysis Techniques; Ontological Models and Applications (two sessions); Heterogeneous Database Integration; Text Analysis in Humanities (two sessions); Data Analysis Projects in Various DID; Organization of Experiments in Data-Intensive Research; Digital Library Projects; Knowledge Representation and Discovery; Approaches for Problem Solving in DID; and Applications of Machine Learning. Although most of the presentations are dedicated to the results of research conducted in organizations in the territory of the Russian Federation including Kazan, Moscow, Novosibirsk, Obninsk, Omsk, Orel, Pereslavl-Zalessky, Saint Petersburg, Tomsk, Yaroslavl, Zvenigorod, the DAMDID/RCDL 2017 conference also had international features. This move is witnessed by 12 talks (four of them are invited) prepared by the notable foreign researchers from such countries as Armenia (Yerevan), Bahrain (Manama), Belarus (Minsk), Bulgaria (Sofia), Germany (Dusseldorf, Kiel), UK (Harvel), Greece (Thessaloniki), Italy (Milan, Naples), and the USA (Harvard).

For the proceedings 19 papers were selected by the Program Committee (16 peer reviewed and three invited papers) and after careful editing they are included in this volume structured into seven sections comprising Data Analytics: two papers; Next-Generation Genomic Sequencing (Challenges and Solutions): two papers; Novel Approaches to Analyzing and Classifying of Various Astronomical Entities and Events: six papers; Ontology Population in Data-Intensive Domains: three papers; Heterogeneous Data Integration Issues: four papers; Data Curation and Data Provenance Support: one paper; Temporal Summaries Generation: one paper. Of these, eight papers (more than one third of the total number of the papers selected) were prepared by foreign researchers (from Bulgaria, Germany, Greece, Italy, UK, USA).

DAMDID/RCDL 2017 would not have been possible without the support of the Russian Foundation for Basic Research, the Federal Agency of Scientific Organizations of the Russian Federation and the Federal Research Center Computer Science and Control of the Russian Academy of Sciences. Finally, we thank Springer for publishing this proceedings volume, containing the invited and selected research papers, in their

CCIS series. The Program Committee of the conference appreciates the possibility to use the Conference Management Toolkit (CMT) sponsored by Microsoft Research, which provided great support during various phases of the paper submission and reviewing process.

May 2018

Leonid Kalinichenko
Yannis Manolopoulos
Oleg Malkov
Nikolay Skvortsov
Sergey Stupnikov
Vladimir Sukhomlin

Organization

General Chair

Igor Sokolov	Federal Research Center Computer Science and Control of RAS, Russia

Program Committee Co-chairs

Leonid Kalinichenko	Federal Research Center Computer Science and Control of RAS, Russia
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece

PhD Workshop Co-chairs

Sergey Stupnikov	Federal Research Center Computer Science and Control of RAS, Russia
Sergey Gerasimov	Lomonosov Moscow State University, Russia

Organizing Committee Co-chairs

Vladimir Sukhomin	Lomonosov Moscow State University, Russia
Victor Zakharov	Federal Research Center Computer Science and Control of RAS, Russia

Organizing Committee

Elena Zubareva	Lomonosov Moscow State University, Russia
Dmitry Briukhov	Federal Research Center Computer Science and Control of RAS, Russia
Nikolay Skvortsov	Federal Research Center Computer Science and Control of RAS, Russia
Dmitry Kovalev	Federal Research Center Computer Science and Control of RAS, Russia
Evgeny Morkovin	Lomonosov Moscow State University, Russia
Irina Karzalova	Federal Research Center Computer Science and Control of RAS, Russia
Yulia Trusova	Federal Research Center Computer Science and Control of RAS, Russia
Evgeniy Ilyushin	Lomonosov Moscow State University, Russia
Dmitry Gouriev	Lomonosov Moscow State University, Russia
Vladimir Romanov	Lomonosov Moscow State University, Russia

Supporters

Russian Foundation for Basic Research
Federal Agency of Scientific Organizations of the Russian Federation
Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences (FRC CSC RAS)
Moscow ACM SIGMOD Chapter

Coordinating Committee

Igor Sokolov (Co-chair)	Federal Research Center Computer Science and Control of RAS, Russia
Nikolay Kolchanov (Co-chair)	Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Leonid Kalinichenko (Deputy Chair)	Federal Research Center Computer Science and Control of RAS, Russia
Arkady Avramenko	Pushchino Radio Astronomy Observatory, RAS, Russia
Pavel Braslavsky	Ural Federal University, SKB Kontur, Russia
Vasily Bunakov	Science and Technology Facilities Council, Harwell, Oxfordshire, UK
Alexander Elizarov	Kazan (Volga Region) Federal University, Russia
Alexander Fazliev	Institute of Atmospheric Optics, RAS, Siberian Branch, Russia
Alexei Klimentov	Brookhaven National Laboratory, USA
Mikhail Kogalovsky	Market Economy Institute, RAS, Russia
Vladimir Korenkov	JINR, Dubna, Russia
Mikhail Kuzminski	Institute of Organic Chemistry, RAS, Russia
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Vladimir Litvine	Evogh Inc., California, USA
Archil Maysuradze	Moscow State University, Russia
Oleg Malkov	Institute of Astronomy, RAS, Russia
Alexander Marchuk	Institute of Informatics Systems, RAS, Siberian Branch, Russia
Igor Nekrestjanov	Verizon Corporation, USA
Boris Novikov	St. Petersburg State University, Russia
Nikolay Podkolodny	ICaG, SB RAS, Novosibirsk, Russia
Aleksey Pozanenko	Space Research Institute, RAS, Russia
Vladimir Serebryakov	Computing Center of RAS, Russia
Yury Smetanin	Russian Foundation for Basic Research, Moscow
Vladimir Smirnov	Yaroslavl State University, Russia
Sergey Stupnikov	Federal Research Center Computer Science and Control of RAS, Russia
Konstantin Vorontsov	Moscow State University, Russia
Viacheslav Wolfengagen	National Research Nuclear University MEPhI, Russia

Victor Zakharov	Federal Research Center Computer Science and Control of RAS, Russia

Program Committee

Karl Aberer	EPFL, Lausanne, Switzerland
Plamen Angelov	Lancaster University, UK
Alexander Afanasyev	Institute for Information Transmission Problems, RAS, Russia
Arkady Avramenko	Pushchino Observatory, Russia
Ladjel Bellatreche	LIAS/ISAE-ENSMA, Poitiers, France
Pavel Braslavski	Ural Federal University, Yekaterinburg, Russia
Vasily Bunakov	Science and Technology Facilities Council, Harwell, UK
Evgeny Burnaev	Skoltech, Russia
George Chernishev	St. Petersburg State University, Russia
Yuri Demchenko	University of Amsterdam, The Netherlands
Boris Dobrov	Research Computing Center of MSU, Russia
Alexander Elizarov	Kazan Federal University, Russia
Alexander Fazliev	Institute of Atmospheric Optics, SB RAS, Russia
Sergey Gerasimov	Lomonosov Moscow State University, Russia
Vladimir Golenkov	Belarusian State University of Informatics and Radioelectronics, Belarus
Vladimir Golovko	Brest State Technical University, Belarus
Olga Gorchinskaya	FORS, Moscow, Russia
Evgeny Gordov	Institute of Monitoring of Climatic and Ecological Systems SB RAS, Russia
Valeriya Gribova	Institute of Automation and Control Processes FEBRAS, Far Eastern Federal University, Russia
Maxim Gubin	Google Inc., USA
Natalia Guliakina	Belarusian State University of Informatics and Radioelectronics, Belarus
Ralf Hofestadt	University of Bielefeld, Germany
Leonid Kalinichenko	FRC CSC RAS, Moscow, Russia
George Karypis	University of Minnesota, Minneapolis, USA
Nadezhda Kiselyova	IMET RAS, Russia
Alexei Klimentov	Brookhaven National Laboratory, USA
Mikhail Kogalovsky	Market Economy Institute, RAS, Russia
Vladimir Korenkov	Joint Institute for Nuclear Research, Dubna, Russia
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Sergei O. Kuznetsov	National Research University Higher School of Economics, Russia
Dmitry Lande	Institute for Information Recording, NASU, Russia
Giuseppe Longo	University of Naples Federico II, Italy
Natalia Loukachevitch	Moscow State University, Russia
Ivan Lukovic	University of Novi Sad, Serbia
Oleg Malkov	Institute of Astronomy, RAS, Russia

Yannis Manolopoulos	School of Informatics of the Aristotle University of Thessaloniki, Greece
Manuel Mazzara	Innopolis University, Russia
Alexey Mitsyuk	National Research University Higher School of Economics, Russia
Xenia Naidenova	S. M. Kirov Military Medical Academy, Russia
Dmitry Namiot	Lomonosov Moscow State University, Russia
Igor Nekrestyanov	Verizon Corporation, USA
Gennady Ososkov	Joint Institute for Nuclear Research, Russia
Dmitry Paley	Yaroslav State University, Russia
Nikolay Podkolodny	Institute of Cytology and Genetics SB RAS, Russia
Natalia Ponomareva	Scientific Center of Neurology of RAMS, Russia
Alexey Pozanenko	Space Research Institute, RAS, Russia
Andreas Rauber	Vienna TU, Austria
Roman Samarev	Bauman Moscow State Technical University, Russia
Timos Sellis	RMIT, Australia
Vladimir Serebryakov	Computing Centre of RAS, Russia
Nikolay Skvortsov	FRC CSC RAS, Russia
Vladimir Smirnov	Yaroslavl State University, Russia
Manfred Sneps-Sneppe	AbavaNet, Russia
Valery Sokolov	Yaroslavl State University, Russia
Sergey Stupnikov	FRC CSC RAS, Russia
Alexander Sychev	Voronezh State University, Russia
Dmitry Tsarkov	Google, USA
Bernhard Thalheim	University of Kiel, Germany
Dmitry Tsarkov	Manchester University, UK
Alexey Ushakov	University of California, Santa Barbara, USA
Natalia Vassilieva	Hewlett-Packard, Russia
Pavel Velikhov	Finstar Financial Group, Russia
Alexey Vovchenko	FRC CSC RAS, Moscow, Russia
Peter Wittenburg	MPI for Psycholinguistics, Germany
Vladimir Zadorozhny	University of Pittsburgh, USA
Yury Zagorulko	Institute of Informatics Systems, SB RAS, Russia
Victor Zakharov	FRC CSC RAS, Russia
Sergey Znamensky	Institute of Program Systems, RAS, Russia

Contents

Data Curation and Data Provenance Support

Temporal Summaries Generation