

Editorial Board Members

Joaquim Filipe 

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Raquel Oliveira Prates 

Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

Lizhu Zhou

Tsinghua University, Beijing, China

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

Alexander Sychev · Sergey Makhortov ·
Bernhard Thalheim (Eds.)

Data Analytics and Management in Data Intensive Domains

22nd International Conference, DAMDID/RCDL 2020
Voronezh, Russia, October 13–16, 2020
Selected Proceedings

Editors

Alexander Sychev
Voronezh State University
Voronezh, Russia

Sergey Makhortov
Voronezh State University
Voronezh, Russia

Bernhard Thalheim 
Christian-Albrecht University of Kiel
Kiel, Schleswig-Holstein, Germany

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-81199-0 ISBN 978-3-030-81200-3 (eBook)
<https://doi.org/10.1007/978-3-030-81200-3>

© Springer Nature Switzerland AG 2021

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, expressed 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 XXII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2020) that was set to be held at the Voronezh State University, Russia, during October 13–16, 2020. However, because of the worldwide COVID-19 crisis, DAMDID/RCDL 2020 had to take place online.

DAMDID is a multidisciplinary forum of researchers and practitioners from various domains of science and research, promoting cooperation and the 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.) and social sciences, as well as in various branches of informatics, industry, new technologies, finance, and business, contribute to the conference content.

Previous DAMDID/RCDL conferences were held in St. Petersburg (1999, 2003), Protvino (2000), Petrozavodsk (2001, 2009), Dubna (2002, 2008, 2014), Pushchino (2004), Yaroslavl (2005, 2013), Pereslavl (2007, 2012), Kazan (2010, 2019), Voronezh (2011), Obninsk (2016), and Moscow (2017, 2018).

The program of DAMDID/RCDL 2020 was oriented towards data science and data-intensive analytics as well as data management topics. The proceedings include abstracts of the two keynotes.

The keynote by Ladjel Bellatreche (full professor at the National Engineering School for Mechanics and Aerotechnics, France) was aimed to motivate researchers to embrace the energy-efficiency of data management systems. Oscar Pastor (full professor and director of the Research Center on Software Production Methods at the Polytechnic University of Valencia, Spain) gave a talk discussing a fundamental role of a sound conceptual modeling in human genome research.

The conference Program Committee, comprised of members from 8 countries, reviewed 60 submissions. 29 submissions were accepted as full papers, 14 as short papers, and 3 as short demos, whilst 14 submissions were rejected.

According to the conference and workshops program, 55 oral presentations were grouped into 11 sessions. Most of the presentations were dedicated to the results of research conducted in Russian organizations including those located in Dubna, Ekaterinburg, Innopolis, Kazan, Krasnodar, Moscow, Novosibirsk, Obninsk, Tomsk, Voronezh, Pereslavl, Perm, St. Petersburg, Petrozavodsk, Tver, Tyumen, and Yaroslavl. However, the conference also featured talks prepared by the foreign researchers from countries such as Armenia, France, Germany, Spain, UK, and USA.

For the CCIS conference proceedings, 16 peer reviewed papers and 2 keynote abstracts have been selected by the Program Committee (acceptance rate: 27%) and-structured into 6 sections including Data Integration, Conceptual Models, and Ontologies (3 papers); Data Management in Semantic Web (3 papers); Data Analysis in

Medicine (2 papers); Data Analysis in Astronomy (3 papers); and Information Extraction from Text (5 papers).

We are grateful to the Program Committee members, for reviewing the submissions and selecting the papers for presentation, to the authors of the submissions, and to the host organizers from Voronezh State University. We are also grateful for the use of the Conference Management Toolkit (CMT) sponsored by Microsoft Research, which provided great support during various phases of the paper submission and reviewing process.

April 2021

Alexander Sychev
Sergey Makhortov
Bernhard Thalheim

Organization

Program Committee Co-chairs

Sergey Makhortov	Voronezh State University, Russia
Alexander Sychev	Voronezh State University, Russia
Bernhard Thalheim	University of Kiel, Germany

Program Committee Deputy Chair

Sergey Stupnikov	Federal Research Center “Computer Science and Control” of RAS, Russia
------------------	---

PhD Workshop Chair

Sergey Makhortov	Voronezh State University, Russia
------------------	-----------------------------------

PhD Workshop Curator

Ivan Lukovic	University of Novi Sad, Serbia
--------------	--------------------------------

Organizing Committee Co-chairs

Oleg Kozaderov	Voronezh State University, Russia
Victor Zakharov	Federal Research Center “Computer Science and Control” of RAS, Russia

Organizing Committee Deputy Chairs

Eduard Algazinov (Deceased)	Voronezh State University, Russia
Sergey Stupnikov	Federal Research Center “Computer Science and Control” of RAS, Russia

Organizing Committee

Andrey Koval	Voronezh State University, Russia
Nikolay Skvortsov	Federal Research Center “Computer Science and Control” of RAS, Russia
Dmitry Borisov	Voronezh State University, Russia
Alexey Vakhtin	Voronezh State University, Russia
Dmitry Briukhov	Federal Research Center “Computer Science and Control” of RAS, Russia
Olga Schepkina	Voronezh State University, Russia

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, Russia
Sergey Stupnikov (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, UK
Alexander Elizarov	Kazan Federal University, Russia
Alexander Fazliev	Institute of Atmospheric Optics, SB RAS, Russia
Alexei Klimentov	Brookhaven National Laboratory, USA
Mikhail Kogalovsky	Market Economy Institute, RAS, Russia
Vladimir Korenkov	Joint Institute for Nuclear Research, Russia
Mikhail Kuzminski	Institute of Organic Chemistry, RAS, Russia
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Vladimir Litvine	Evogh Inc., USA
Archil Maysuradze	Moscow State University, Russia
Oleg Malkov	Institute of Astronomy, RAS, Russia
Alexander Marchuk	Institute of Informatics Systems, SB RAS, Russia
Igor Nekrestjanov	Verizon Corporation, USA
Boris Novikov	Saint Petersburg State University, Russia
Nikolay Podkolodny	Institute of Cytology and Genetics, SB RAS, 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
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

Ladjet Bellatreche	National Engineering School for Mechanics and Aerotechnics, France
Dmitry Borisenkov	Relex, Russia
Pavel Braslavski	Ural Federal University, Russia
Vasily Bunakov	Science and Technology Facilities Council, UK
George Chernishev	Saint Petersburg State University, Russia
Boris Dobrov	Lomonosov Moscow State University, Russia
Alexander Elizarov	Kazan Federal University, Russia

Alexander Fazliev	Institute of Atmospheric Optics, SB RAS, Russia
Yuriy Gapanyuk	Bauman Moscow State Technical University, Russia
Veronika Garshina	Voronezh State University, Russia
Evgeny Gordov	Institute of Monitoring of Climatic and Ecological Systems, SB RAS, Russia
Valeriya Gribova	Far Eastern Federal University, Russia
Maxim Gubin	Google Inc., USA
Sergio Ilarri	University of Zaragoza, Spain
Mirjana Ivanovic	University of Novi Sad, Serbia
Nadezhda Kiselyova	IMET RAS, Russia
Vladimir Korenkov	Joint Institute for Nuclear Research, Russia
Sergey Kuznetsov	Institute for System Programming, RAS, Russia
Evgeny Lipachev	Kazan Federal University, Russia
Natalia Loukachevitch	Lomonosov Moscow State University, Russia
Ivan Lukovic	University of Novi Sad, Serbia
Oleg Malkov	Institute of Astronomy, RAS, Russia
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece
Archil Maysuradze	Lomonosov Moscow State University, Russia
Manuel Mazzara	Innopolis University, Russia
Alexey Mitsyuk	National Research University Higher School of Economics, Russia
Xenia Naidenova	Kirov Military Medical Academy, Russia
Dmitry Namiot	Lomonosov Moscow State University, Russia
Dmitry Nikitenko	Lomonosov Moscow State University, Russia
Panos Pardalos	University of Florida, USA
Natalya Ponomareva	Research Center of Neurology, Russia
Alexey Pozanenko	Space Research Institute, RAS, Russia
Roman Samarev	Bauman Moscow State Technical University, Russia
Timos Sellis	Swinburne University of Technology, Australia
Vladimir Serebryakov	Computing Centre of RAS, Russia
Nikolay Skvortsov	Federal Research Center “Computer Science and Control” of RAS, Russia
Manfred Sneps-Sneppe	AbavaNet, Russia
Sergey Sobolev	Lomonosov Moscow State University, Russia
Valery Sokolov	Yaroslavl State University, Russia
Alexey Ushakov	University of California, Santa Barbara, USA
Pavel Velikhov	Huawei, Russia
Alexey Vovchenko	Federal Research Center “Computer Science and Control” of RAS, Russia
Vladimir Zadorozhny	University of Pittsburgh, USA
Yury Zagorulko	Institute of Informatics Systems, SB RAS, Russia
Victor Zakharov	Federal Research Center “Computer Science and Control” of RAS, Russia
Sergey Znamensky	Institute of Program Systems, RAS, Russia
Mikhail Zymbler	South Ural State University, Russia

Supporters

Voronezh State University

Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia

Moscow ACM SIGMOD Chapter

Keynote Speakers' Bios

Ladjel Bellatreche (National Engineering School for Mechanics and Aerotechnics of Aerospace Engineering Group, France)

Ladjel Bellatreche is a Full Professor at National Engineering School for Mechanics and Aerotechnics (ENSMA) of Aerospace Engineering Group (ISAE), Poitiers, France, where he joined as a Faculty Member since September 2010. He leads the Data Engineering Group of the Laboratory of Computer Science & Automatic Control for Systems (LIAS). He was Director of the LISI laboratory from January 2011 till January 2012. Prior to that, he spent eight years as an Assistant and then an Associate Professor at the Poitiers University, France. Ladjel was a Visiting Professor at the Harbin Institute of Technology (HIT), China (2018, 2019), the Québec en Outaouais, Canada (2009), a Visiting Researcher at the Purdue University, USA (2001), and the Hong Kong University of Science and Technology, China (1997–1999).

His current research interests include the different phases of the life cycle of designing data management systems/applications: requirement engineering, conceptual modelling using ontologies and knowledge graphs, logical modelling, Extraction-Transformation-Loading, deployment, physical design, personalisation and recommendation. Ladjel (co)-supervised 30 Ph.D. students on different topics covering the above phases.

Óscar Pastor (Polytechnic University of Valencia, Spain)

Óscar Pastor is Full Professor and Director of the “Software Production Methods” (PROS) at the Polytechnic University of Valencia (Spain). He received his Ph.D. in 1992. Supervisor of 20 completed PhD theses and 31 completed Masters theses on topics that relate to Conceptual Modeling, he has published more than three hundred research papers in conference proceedings, journals and books, received numerous research grants from public institutions and private industry, and been keynote speaker at several conferences and workshops.

He is Chair of the ER Steering Committee (2009–10), ER Fellow since 2010, member of the SC of conferences as CAiSE, ER, ICWE, ESEM, CIbSE or RCIS, and member of over 100 Scientific Committees of top-ranked international conferences, his research activities focus on conceptual modeling, web engineering, requirements engineering, information systems, and model-based software production.

Óscar Pastor created the object-oriented, formal specification language OASIS and the corresponding software production method OO-METHOD. He led the research and development underlying CARE Technologies that was formed in 1996. CARE Technologies has created an advanced MDA-based Conceptual Model Compiler called Integranova, a tool that produces a final software product starting from a conceptual schema that represents system requirements. He is currently leading a multidisciplinary project linking Information Systems and Bioinformatics notions, oriented to designing and implementing tools for Conceptual Modeling-based interpretation of the Human Genome information.

Keynote Abstracts

Towards Green Data Management Systems

Ladjel Bellatreche 

LIAS/ISAE-ENSMA, Poitiers University,
Futuroscope, 86960 Poitiers, France
`ladjel.bellatreche@ensma.fr`

Abstract. In today's world, our life depends too much on computers. Therefore, we are forced to look at every way to save the energy of our hardware components, system software, as well as applications. Data Management Systems (DMSYSs) are at the heart of the energy new world order. The query processor is one of the DMSYS components in charge of the efficient processing of data. Studying the Energy-Efficiency of this component has become an urgent necessity. Most query optimizers minimize inputs/outputs operations and try to exploit RAM as much as possible. Unfortunately, they generally ignore energy aspects. Furthermore, many researchers have the opinion that only the OS and firmware that should manage energy, leaving DMSYSs as a second priority. In our opinion, software and hardware solutions must be integrated to maximize energy savings. This integration seems natural since query optimizers use cost models to select the best query plans and use hardware and software parameters. As scientists, we first feel obliged to motivate researchers to embrace the Energy-Efficiency of DMSYSs through a survey. Secondly, to accompany them, we propose a road-map covering the recent hardware and software solutions impacting query processors. Finally, guidelines for developing green query optimizers are widely discussed.

Conceptual Modeling and Life Engineering: Facing Data Intensive Domains Under a Common Perspective

Oscar Pastor 

PROS Research Center, Universitat Politècnica de València,
Camino Vera s/n., 46022 Valencia, Spain
opastor@pros.upv.es

Abstract. Understanding the Human Genome is the big scientific challenge of our century. It will probably lead to a new kind of “Homo Sapiens” with new capabilities never before affordable for human being as we know them. This will be referred in the keynote as an “Homo Genius” evolution. Getting a shared understanding of the domain is a first essential task in order to manage correctly and efficiently such a complex data intensive domain. With more and more data being generated day after day with the continuous improvements of sequencing technologies, selecting the right data management strategy intended to support the design of the right software platforms that successfully attend user requirements (in terms of relevant information needs) becomes a unavoidable, crucial goal. This talk discussed how the use of a sound conceptual modeling support is fundamental in order to convert that “understanding the genome challenge” into a Life Engineering problem, where conceptual modeling, explainable artificial intelligence and data science must work together in order to provide accurate, reliable and valuable solutions, putting an special emphasis in the modern Medicine of Precision applications.

Contents

Data Integration, Conceptual Models and Ontologies

Managing Data-Intensive Research Problem-Solving Lifecycle	3
<i>Nikolay Skvortsov and Sergey Stupnikov</i>	
Algebraic Models for Big Data and Knowledge Management	19
<i>Sergey Makhortov</i>	
A Cloud-Native Serverless Approach for Implementation of Batch Extract-Load Processes in Data Lakes	27
<i>Anton Bryzgalov and Sergey Stupnikov</i>	

Data Management in Semantic Web

Pragmatic Interoperability and Translation of Industrial Engineering Problems into Modelling and Simulation Solutions	45
<i>Martin T. Horsch, Silvia Chiacchiera, Michael A. Seaton, Ilian T. Todorov, Björn Schembera, Peter Klein, and Natalia A. Konchakova</i>	
Analysis of the Semantic Distance of Words in the RuWordNet Thesaurus	60
<i>Liliya Usmanova, Irina Erofeeva, Valery Solovyev, and Vladimir Bochkarev</i>	
A Transformation of the RDF Mapping Language into a High-Level Data Analysis Language for Execution in a Distributed Computing Environment	74
<i>Wenfei Tang and Sergey Stupnikov</i>	

Data Analysis in Medicine

EMG and EEG Pattern Analysis for Monitoring Human Cognitive Activity during Emotional Stimulation	95
<i>Konstantin Sidorov, Natalya Bodrina, and Natalya Filatova</i>	
Finding the TMS-Targeted Group of Fibers Reconstructed from Diffusion MRI Data.	110
<i>Sofya Kulikova and Aleksey Buzmakov</i>	

Data Analysis in Astronomy

Data for Binary Stars from Gaia DR2 125
*Dana Kovaleva, Oleg Malkov, Sergei Sapozhnikov, Dmitry Chulkov,
and Nikolay Skvortsov*

Classification Problem and Parameter Estimating of Gamma-Ray Bursts 134
Pavel Minaev and Alexei Pozanenko

Databases of Gamma-Ray Bursts’ Optical Observations 148
*Alina Volnova, Alexei Pozanenko, Elena Mazaeva, Sergei Belkin,
and Pavel Minaev*

Information Extraction from Text

Part of Speech and Gramset Tagging Algorithms for Unknown
Words Based on Morphological Dictionaries of the Veps
and Karelian Languages. 163
Andrew Krizhanovsky, Natalia Krizhanovskaya, and Irina Novak

Extrinsic Evaluation of Cross-Lingual Embeddings on the Patent
Classification Task 178
Anastasiia Ryzhova and Ilya Sochenkov

An Approach to Extracting Ontology Concepts from Requirements. 191
Marina Murtazina and Tatiana Avdeenko

Data Driven Detection of Technological Trajectories 204
Sergey Volkov, Dmitry Devyatkin, Ilya Tikhomirov, and Ilya Sochenkov

Comparison of Cross-Lingual Similar Documents Retrieval Methods. 216
Denis Zubarev and Ilya Sochenkov

Author Index 231